(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 662 133 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
18.11.2009 Bulletin 2009/47

(21) Application number: **93922703.9**

(22) Date of filing: **17.09.1993**

(51) Int Cl.:
C12N 15/51 *(2006.01)* ***C12N 7/00*** *(2006.01)*
C12N 15/40 *(2006.01)* ***A61K 39/29*** *(2006.01)*
G01N 33/576 *(2006.01)*

(86) International application number:
PCT/US1993/008849

(87) International publication number:
WO 1994/006913 (31.03.1994 Gazette 1994/08)

(54) **RECOMBINANT PROTEINS OF A PAKISTANI STRAIN OF HEPATITIS E AND THEIR USE IN DIAGNOSTIC METHODS AND VACCINES**

REKOMBINANTE PROTEINE EINES PAKISTANI STAMMES VON HEPATITIS E-VIRUS SOWIE IHRE VERWENDUNG IN DIAGNOSTISCHEN VERFAHREN UND ALS VAKZINE

PROTEINES RECOMBINEES D'UNE SOUCHE PAKISTANAISE DU VIRUS DE L'HEPATITE E ET LEUR UTILISATION DANS DES PROCEDES DIAGNOSTIQUES ET DES VACCINS

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **18.09.1992 US 947263**

(43) Date of publication of application:
12.07.1995 Bulletin 1995/28

(73) Proprietor: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES
Rockville, MD 20852 (US)**

(72) Inventors:
- **TSAREV, Sergei, A.
Bethesda, MD 20814 (US)**
- **EMERSON, Suzanne, U.
Rockville, MD 20853 (US)**
- **PURCELL, Robert, H.
Boyds, MD 20841 (US)**

(74) Representative: **Vossius & Partner
Siebertstraße 4
81675 München (DE)**

(56) References cited:
WO-A-93/14208

- **PROC NAT'L. ACAD. SCI. USA vol. 89 , 1992 pages 559 - 563 S. TSAREV ET AL. 'Characterisation of a prototype strain of hepatitis E virus' & EMBL Database entry HPEA Accession number M80581, 24 February 1992**
- **ARCH. VIROL. vol. 123 , 1992 pages 335 - 349 M. PURDY ET AL. 'Expression of a hepatitis E virus-trpE fusion protein'**
- **VIROLOGY vol. 185 , 1991 pages 120 - 131 A. TAM ET AL. 'Hepatitis E virus: molecular cloning and sequencing of the full-length viral genome'**
- **J. GEN. VIROL. vol. 73 , 1992 pages 1143 - 1148 R. HUANG ET AL. 'Isolation and identification of hepatitis E virus in China'**
- **J. INFECT. DISEASES vol. 168, no. 2 , 1993 pages 369 - 378 S. TSAREV ET AL. 'ELISA for antibody to heaptitis E virus based on complete open reading frame 2 expressed in onsect cells'**
- **J. CLIN. MICROBIOL. vol. 31, no. 8 , 1993 pages 2167 - 2173 J. HE AT AL. 'Expression and diagnostic utility of hepatitis E virus putative structural proteins expressed in insect cells'**
- **PURCELL R.H.: 'Viral Hepatits and liver disease: Proceedings of the 10th international symposium on viral hepatitis and liver disease', 2002, INTERNATIONAL MEDICAL PRESS LTD, ATLANTA * page 97 - page 102 ***

Remarks:
The file contains technical information submitted after the application was filed and not included in this specification

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 0 662 133 B1

Description

[0001] The invention is in the field of hepatitis virology. More specifically, this invention relates to a method for producing an immunogenic hepatitis protein, namely to the expression of a complete open reading frame 2 coding sequence of a hepatitis E virus in an insect host cell transformed or transfected with a baculovirus expression vector comprising said coding sequence, to the baculovirus expression vector and the insect host cell.

[0002] Epidemics of hepatitis E, an enterically transmitted non-A/non-B hepatitis, have been reported in Asia, Africa and Central America (Balayan, M.S. (1987), Soviet Medical Reviews, Section E, Virology Reviews, Zhdanov, V.M. (ed), Chur, Switzerland: Harwood Academic Publishers, vol. 2, 235-261; Purcell, R.H., et al. (1988) in Zuckerman, A.J. (ed), "Viral Hepatitis and Liver Disease", New York: Alan R. Liss, 131-137; Bradley, D.W. (1990), British Medical Bulletin, 46: 442-461; Ticehurst, J.R. (1991) in Hollinger, F.B., Lemon, S.M., Margolis, H.S. (eds): "Viral Hepatitis and Liver Disease", Williams and Wilkins, Baltimore, 501-513). Cases of sporadic hepatitis, presumed to be hepatitis E, account for up to 90% of reported hepatitis in countries where hepatitis E virus (HEV) is endemic. The need for development of a serological test for the detection of anti-HEV antibodies in the sera of infected individuals is widely recognized in the field, but the very low concentration of HEV excreted from infected humans or animals made it impossible to use such HEV as the source of antigen for serological tests and although limited success was reported in propagation of HEV in cell culture (Huang, R.T. et al. (1992), J. Gen. Virol., 73:1143-1148), cell culture is currently too inefficient to produce the amounts of antigen required for serological tests.

[0003] Recently, major efforts worldwide to identify viral genomic sequences associated with hepatitis E have resulted in the cloning of the genomes of a limited number of strains of HEV (Tam, A.W. et al. (1991), Virology, 185:120-131; Tsarev, S.A. et al. (1992), Proc, Natl, Acad. Sci. USA, 89:559-563; Fry, K.E. et al. (1992), Virus Genes, 6:173-185). Analysis of the DNA sequences have led investigators to hypothesize that the HEV genome is organized into three open reading frames (ORPs) and to hypothesize that these ORFs encode intact HEV proteins.

[0004] A partial DNA sequence of the genome of an HEV strain from Burma (Myanmar) is disclosed in Reyes et al. 1990, Science, 247:1335-1339. Tam et al., 1991, and Reyes et al., PCT Patent Application WO91/15603 published October 17, 1991 disclose the complete nucleotide sequence and a deduced amino acid sequence of the Burma strain of HEV. These authors hypothesized that three forward open reading frames (ORFs) are contained within the sequence of this strain.

[0005] Ichikawa et al., 1991, Microbiol. Immunol., 35:535-543, discloses the isolation of a series of clones of 240-320 nucleotides in length upon the screening of a λgt11 expression library with sera from HEV-infected cynomolgus monkeys. The recombinant protein expressed by one clone was expressed in E. coli. This fusion protein is encoded by the 3' region of ORF-2 of the Myanmar strain of HEV.

[0006] The expression of additional proteins encoded within the 3' region of ORF-2 of a Mexican strain of HEV and of a Burmese strain of HEV is described in Yarbough et al., 1991 J. Virology, 65:5790-5797. This article describes the isolation of two cDNA clones derived from HEV. These clones encode the proteins in the 3' region of ORF-2. The clones were expressed in E. coli as fusion proteins.

[0007] Purdy et al., 1992, Archives of Virology, 123:335-349, and Favorov et al., 1992, J. of Medical Virology, 36: 246-250, disclose the expression of a larger ORF-2 protein fragment from the Burma strain in E. coli. These references, as well as those previously discussed, only disclose the expression of a portion of the ORF-2 gene using bacterial expression systems. Successful expression of the full-length ORF-2 protein has not been disclosed until the present invention.

[0008] Comparison of the genome organization and morphological structure of HEV to that of other viruses revealed that HEV is most closely related to the caliciviruses. Of interest, the structural proteins of caliciviruses are encoded by the 3' portion of their genome (Neil, J.D. et al. (1991) J. Virol., 65:5440-5447; and Carter, M.J. et al. (1992), J. Arch. Virol., 122:223-235) and although there is no direct evidence that the 3' terminal part of the HEV genome also encodes the structural proteins, expression of certain small portions of the 3' genome region in bacterial cells resulted in production of proteins reactive with anti-HEV sera in ELISA and Western blots (Yarborough, et al., (1991); Ichikawa et al. (1991); Favorov et al. (1992) and Dawson, G.J. et al. (1992) J. Virol. Meth; 38:175-186). However, the function of ORF-2 protein as a structural protein was not proven until the present invention.

[0009] The small proteins encoded by a portion of the ORF-2 gene have been used in immunoassay to detect antibodies to HEV in animal sera. The use of small bacterially expressed proteins as antigens in serological immunoassays has several potential drawbacks. First, the expression of these small proteins in bacterial cells often results in solubility problems and in non-specific cross-reactivity of patients' sera with E. coli proteins when crude E. coli lysates are used as antigens in immunoassays (Purdy et al. (1992)). Second, the use of Western blots as a first-line serological test for anti-HEV antibodies in routine epidemiology is impractical due to time and cost constraints. An ELISA using small-peptides derived from the 3'-terminal part of the HEV genome resulted in the detection of only 41% positives from known HEV-infected patients. Third, it has been shown that for many viruses, including *Picornaviridae* which is the closest family to the *Caliciviridae*, important antigenic and immunogenic epitopes are highly conformational (Lemon, S.M. et al.

(1991), in Hollinger, F.B., Lemon, S.M., Margolis, H.S. (eds): "Viral Hepatitis and Liver Disease", Williams and Wilkins, Baltimore, 20-24). For this reason, it is believed that expression in a eukaryotic system of a complete ORF encoding an intact HEV gene would result in production of a protein which could form HEV-virus-like particles. Such a complete ORF protein would have an immunological structure closer to that of native capsid protein(s) than would the above-noted smaller proteins which represent only portions of the structural proteins of HEV. Therefore, these complete ORF proteins would likely serve as a more representative antigen and a more efficient immunogen than the currently-used smaller proteins.

[0010] The present invention relates to a:

method for producing an immunogenic hepatitis protein comprising:

culturing an insect host cell transformed or transfected with a baculovirus expression vector, said expression vector comprising DNA, said DNA having a sequence, said sequence consisting of the nucleotides which encode the amino acids 1 to 660 of a hepatitis E virus open reading frame 2 protein, under conditions appropriate to cause production of the immunogenic hepatitis protein.

[0011] Preferably, the present invention relates to the above method, wherein said DNA sequence consists of:

a) nucleotides 5147 to 7126 of SEQ ID NO: 4,
b) a DNA sequence which encodes the amino acid sequence of SEQ ID NO: 2 or
c) a DNA sequence which encodes an amino acid sequence that is in excess of 90% homologous with the amino acid sequence of SEQ ID NO: 2.

[0012] Moreover, the present invention relates to a baculovirus expression vector comprising DNA, said DNA having a sequence, said sequence consisting of the nucleotides which encode the amino acids 1 to 660 of a hepatitis E virus open reading frame 2 protein.

[0013] Preferably the baculovirus expression vector is an expression vector, wherein said DNA sequence consists of:

a) nucleotides 5147 to 7126 of SEQ ID NO: 4,
b) a DNA sequence which encodes the amino acid sequence of SEQ ID NO: 2 , or
c) a DNA sequence which encodes an amino acid sequence that is in excess of 90% homologous with the amino acid sequence of SEQ ID NO: 2.

[0014] The present invention also relates to an insect host cell comprising the expression vector mentioned above.

[0015] This invention provides the use of synthetic nucleic acid sequences capable of directing production of recombinant HEV proteins, as well as equivalent natural nucleic acid sequences. Such natural nucleic acid sequences maybe isolated from a cDNA or genomic library from which the gene capable of directing synthesis of the HEV proteins may be identified and isolated. For purpose of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any synthetic variant thereof which encodes for protein.

[0016] The invention also relates to isolated and substantially purified HEV proteins and variants thereof encoded by the HEV genome of SAR-55 or encoded by synthetic nucleic acid sequences and in particular to recombinant proteins encoded by at least one complete open reading frame of HEV, as a result of the method of the present invention.

[0017] The invention also relates to the method of preparing recombinant HEV proteins derived from an HEV genomic sequence by cloning the nucleic acid and inserting the cDNA into an expression vector and expressing the recombinant protein in a host cell within the scope of the method of the present invention

Figure 1 shows the recombinant vector used to express the complete ORF-2 protein of the genome of HEV strain SAR-55.

Figures 2A and 2B are sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) in which cell lysates of insect cells infected with wild-type baculovirus or recombinant baculovirus (containing the gene encoding ORF-2) were either stained with Coomassie blue (A) or subjected to Western blotting with serum of an HEV-infected chimp (B).

Figures 3A and 3B show immunoelectron micrographs (IEM) of 30 and 20 nm virus-like particles respectively, which are formed as a result of the expression of ORF-2 protein in recombinant infected insect cells.

Figure 4 shows the results of an ELISA using as the antigen, recombinant ORF-2 which was expressed from insect cells containing the gene encoding the complete ORF-2. Serum anti-HEV antibody levels were determined at various times following inoculation of cynomolgus monkeys with either the Mexican (Cyno-80A82, Cyno-9A97 and Cyno 83) or Pakistani (Cyno-374) strains of HEV.

Figures 5A-D show the results of an ELISA using as the antigen, recombinant ORF-2 which was expressed from

insect cells containing the gene encoding the complete ORF-2. Serum IgG or IgM anti-HEV levels were determined over time following inoculation of two chimpanzees with HEV.
Figures 6A-J show a comparison of ELISA data obtained using as the antigen the recombinant complete ORF-2 protein derived from SAR-55 as the antigen vs. a recombinant partial ORF-2 protein derived from the Burma strain of HEV (Genelabs).

[0018] The present invention relates to a method for producing an immunogenic hepatitis protein comprising:

culturing an insect host cell transformed or transfected with a baculovirus expression vector, said expression vector comprising DNA, said DNA having a sequence, said sequence consisting of the nucleotides which encode the amino acids 1 to 660 of a hepatitis E virus open reading frame 2 protein, under conditions appropriate to cause production of the immunogenic hepatitis protein.

[0019] The present invention discloses an isolated and substantially purified strain of hepatitis E virus (HEV) from Pakistan, SAR-55. The present invention also discloses to the cloning of the viral genes encoding proteins of HEV and the expression of the recombinant proteins using an expression system. More specifically, the present invention discloses the cloning and expression of the open reading frames (ORF) of HEV derived from SAR-55.
[0020] The present invention discloses isolated HEV proteins. Preferably, the HEV proteins as a result of the method of the present invention are substantially homologous to, and most preferably biologically equivalent to, the native HEV proteins. By "biologically equivalent" as used throughout the specification and claims, it is meant that the compositions are capable of forming viral-like particles and are immunogenic. The HEV proteins of the present invention may also stimulate the production of protective antibodies upon injection into a mammal that would serve to protect the mammal upon challenge with a wild type HEV. By "substantially homologous" as used throughout the ensuing specification and claims, is meant a degree of homology in the amino acid sequence to the native HEV proteins. Preferably the degree of homology is in excess of 70%, preferably in excess of 90%, with a particularly preferred group of proteins being in excess of 99% homologous with the native HEV proteins.
[0021] Preferred HEV proteins produced by the method of the present invention are those proteins that are encoded by the ORF genes including the ORF-2 gene. Of particular interest are proteins encoded by the ORF-2 gene of HEV and most particularly proteins encoded by the ORF-2 gene of the SAR-55 strain of HEV. The proteins produced by the method of the present invention, encoded by the ORF-2 gene, form virus-like particles. The amino acid sequences of the ORF-1, ORF-2 and ORF-3 proteins are shown below as SEQ ID NO.: 1, SEQ ID NO.: 2, and SEQ ID NO.: 3, respectively:

(SEQ. ID NO.: 1)

```
Met Glu Ala His Gln Phe Ile Lys Ala Pro Gly Ile Thr Thr Ala
1               5                  10                  15

Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser Ala Leu Ala Asn
            20                  25                  30

Ala Val Val Val Arg Pro Phe Leu Ser His Gln Gln Ile Glu Ile
        35                  40                  45

Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg Pro Glu
    50                  55                  60

Val Phe Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu Leu
65                  70                  75

Glu Leu Tyr Cys Arg Ala Arg Ser Gly Arg Cys Leu Glu Ile Gly
                80                  85                  90

Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Val His Arg
            95                  100                 105

Cys Phe Leu Arg Pro Ala Gly Arg Asp Val Gln Arg Trp Tyr Thr
        110                 115                 120

Ala Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu
    125                 130                 135

Arg Gly Leu Pro Ala Ala Asp Arg Thr Tyr Cys Phe Asp Gly Phe
140                 145                 150

Ser Gly Cys Asn Phe Pro Ala Glu Thr Gly Ile Ala Leu Tyr Ser
155                 160                 165

Leu His Asp Met Ser Pro Ser Asp Val Ala Glu Ala Met Phe Arg
            170                 175                 180

His Gly Met Thr Arg Leu Tyr Ala Ala Leu His Leu Pro Pro Glu
        185                 190                 195

Val Leu Leu Pro Pro Gly Thr Tyr Arg Thr Ala Ser Tyr Leu Leu
    200                 205                 210

Ile His Asp Gly Arg Arg Val Val Val Thr Tyr Glu Gly Asp Thr
215                 220                 225

Ser Ala Gly Tyr Asn His Asp Val Ser Asn Leu Arg Ser Trp Ile
            230                 235                 240

Arg Thr Thr Lys Val Thr Gly Asp His Pro Leu Val Ile Glu Arg
        245                 250                 255

Val Arg Ala Ile Gly Cys His Phe Val Leu Leu Leu Thr Ala Ala
    260                 265                 270

Pro Glu Pro Ser Pro Met Pro Tyr Val Pro Tyr Pro Arg Ser Thr
275                 280                 285

Glu Val Tyr Val Arg Ser Ile Phe Gly Pro Gly Gly Thr Pro Ser
            290                 295                 300

Leu Phe Pro Thr Ser Cys Ser Thr Lys Ser Thr Phe His Ala Val
305                 310                 315
```

Pro Ala His Ile Trp Asp Arg Leu Met Leu Phe Gly Ala Thr Leu
320 325 330

Asp Asp Gln Ala Phe Cys Cys Ser Arg Leu Met Thr Tyr Leu Arg
335 340 345

Gly Ile Ser Tyr Lys Val Thr Val Gly Thr Leu Val Ala Asn Glu
350 355 360

Gly Trp Asn Ala Ser Glu Asp Ala Leu Thr Ala Val Ile Thr Ala
365 370 375

Ala Tyr Leu Thr Ile Cys His Gln Arg Tyr Leu Arg Thr Gln Ala
380 385 390

Ile Ser Lys Gly Met Arg Arg Leu Glu Arg Glu His Ala Gln Lys
395 400 405

Phe Ile Thr Arg Leu Tyr Ser Trp Leu Phe Glu Lys Ser Gly Arg
410 415 420

Asp Tyr Ile Pro Gly Arg Gln Leu Glu Phe Tyr Ala Gln Cys Arg
425 430 435

Arg Trp Leu Ser Ala Gly Phe His Leu Asp Pro Arg Val Leu Val
440 445 450

Phe Asp Glu Ser Ala Pro Cys His Cys Arg Thr Ala Ile Arg Lys
455 460 465

Ala Val Ser Lys Phe Cys Cys Phe Met Lys Trp Leu Gly Gln Glu
470 475 480

Cys Thr Cys Phe Leu Gln Pro Ala Glu Gly Val Val Gly Asp Gln
485 490 495

Gly His Asp Asn Glu Ala Tyr Glu Gly Ser Asp Val Asp Pro Ala
500 505 510

Glu Ser Ala Ile Ser Asp Ile Ser Gly Ser Tyr Val Val Pro Gly
515 520 525

Thr Ala Leu Gln Pro Leu Tyr Gln Ala Leu Asp Leu Pro Ala Glu
530 535 540

Ile Val Ala Arg Ala Gly Arg Leu Thr Ala Thr Val Lys Val Ser
545 550 555

Gln Val Asp Gly Arg Ile Asp Cys Glu Thr Leu Leu Gly Asn Lys
560 565 570

Thr Phe Arg Thr Ser Phe Val Asp Gly Ala Val Leu Glu Thr Asn
575 580 585

Gly Pro Glu Arg His Asn Leu Ser Phe Asp Ala Ser Gln Ser Thr
590 595 600

Met Ala Ala Gly Pro Phe Ser Leu Thr Tyr Ala Ala Ser Ala Ala
605 610 615

Gly Leu Glu Val Arg Tyr Val Ala Ala Gly Leu Asp His Arg Ala
620 625 630

Val Phe Ala Pro Gly Val Ser Pro Arg Ser Ala Pro Gly Glu Val
635 640 645

Thr Ala Phe Cys Ser Ala Leu Tyr Arg Phe Asn Arg Glu Ala Gln
650 655 660

Arg Leu Ser Leu Thr Gly Asn Phe Trp Phe His Pro Glu Gly Leu
665 670 675

Leu Gly Pro Phe Ala Pro Phe Ser Pro Gly His Val Trp Glu Ser
680 685 690

Ala Asn Pro Phe Cys Gly Glu Ser Thr Leu Tyr Thr Arg Thr Trp
695 700 705

Ser Glu Val Asp Ala Val Pro Ser Pro Ala Gln Pro Asp Leu Gly
710 715 720

Phe Thr Ser Glu Pro Ser Ile Pro Ser Arg Ala Ala Thr Pro Thr
725 730 735

Pro Ala Ala Pro Leu Pro Pro Pro Ala Pro Asp Pro Ser Pro Thr
740 745 750

Leu Ser Ala Pro Ala Arg Gly Glu Pro Ala Pro Gly Ala Thr Ala
755 760 765

Arg Ala Pro Ala Ile Thr His Gln Thr Ala Arg His Arg Arg Leu
770 775 780

Leu Phe Thr Tyr Pro Asp Gly Ser Lys Val Phe Ala Gly Ser Leu
785 790 795

Phe Glu Ser Thr Cys Thr Trp Leu Val Asn Ala Ser Asn Val Asp
800 805 810

His Arg Pro Gly Gly Gly Leu Cys His Ala Phe Tyr Gln Arg Tyr
815 820 825

Pro Ala Ser Phe Asp Ala Ala Ser Phe Val Met Arg Asp Gly Ala
830 835 840

Ala Ala Tyr Thr Leu Thr Pro Arg Pro Ile Ile His Ala Val Ala
845 850 855

Pro Asp Tyr Arg Leu Glu His Asn Pro Lys Arg Leu Glu Ala Ala
860 865 870

Tyr Arg Glu Thr Cys Ser Arg Leu Gly Thr Ala Ala Tyr Pro Leu
875 880 885

Leu Gly Thr Gly Ile Tyr Gln Val Pro Ile Gly Pro Ser Phe Asp
890 895 900

Ala Trp Glu Arg Asn His Arg Pro Gly Asp Glu Leu Tyr Leu Pro
905 910 915

Glu Leu Ala Ala Arg Trp Phe Glu Ala Asn Arg Pro Thr Cys Pro
920 925 930

Thr Leu Thr Ile Thr Glu Asp Val Ala Arg Thr Ala Asn Leu Ala
935 940 945

Ile Glu Leu Asp Ser Ala Thr Asp Val Gly Arg Ala Cys Ala Gly
950 955 960

Cys Arg Val Thr Pro Gly Val Val Gln Tyr Gln Phe Thr Ala Gly
965 970 975

Val Pro Gly Ser Gly Lys Ser Arg Ser Ile Thr Gln Ala Asp Val
980 985 990

Asp Val Val Val Val Pro Thr Arg Glu Leu Arg Asn Ala Trp Arg
995 1000 1005

Arg Arg Gly Phe Ala Ala Phe Thr Pro His Thr Ala Ala Arg Val
1010 1015 1020

Thr Gln Gly Arg Arg Val Val Ile Asp Glu Ala Pro Ser Leu Pro
1025 1030 1035

Pro His Leu Leu Leu Leu His Met Gln Arg Ala Ala Thr Val His
1040 1045 1050

Leu Leu Gly Asp Pro Asn Gln Ile Pro Ala Ile Asp Phe Glu His
1055 1060 1065

Ala Gly Leu Val Pro Ala Ile Arg Pro Asp Leu Ala Pro Thr Ser
1070 1075 1080

Trp Trp His Val Thr His Arg Cys Pro Ala Asp Val Cys Glu Leu
1085 1090 1095

Ile Arg Gly Ala Tyr Pro Met Ile Gln Thr Thr Ser Arg Val Leu
1100 1105 1110

Arg Ser Leu Phe Trp Gly Glu Pro Ala Val Gly Gln Lys Leu Val
1115 1120 1125

Phe Thr Gln Ala Ala Lys Ala Ala Asn Pro Gly Ser Val Thr Val
1130 1135 1140

His Glu Ala Gln Gly Ala Thr Tyr Thr Glu Thr Thr Ile Ile Ala
1145 1150 1155

Thr Ala Asp Ala Arg Gly Leu Ile Gln Ser Ser Arg Ala His Ala
1160 1165 1170

Ile Val Ala Leu Thr Arg His Thr Glu Lys Cys Val Ile Ile Asp
1175 1180 1185

Ala Pro Gly Leu Leu Arg Glu Val Gly Ile Ser Asp Ala Ile Val
1190 1195 1200

Asn Asn Phe Phe Leu Ala Gly Gly Glu Ile Gly His Gln Arg Pro
1205 1210 1215

Ser Val Ile Pro Arg Gly Asn Pro Asp Ala Asn Val Asp Thr Leu
1220 1225 1230

Ala Ala Phe Pro Pro Ser Cys Glu Ile Ser Ala Phe His Glu Leu
1235 1240 1245

Ala Glu Glu Leu Gly His Arg Pro Ala Pro Val Ala Ala Val Leu
1250 1255 1260

Pro Pro Cys Pro Glu Leu Glu Gln Gly Leu Leu Tyr Leu Pro Gln
1265 1270 1275

Glu Leu Thr Thr Cys Asp Ser Val Val Thr Phe Glu Leu Thr Asp
1280 1285 1290

Ile Val His Cys Arg Met Ala Ala Pro Ser Gln Arg Lys Ala Val
1295 1300 1305

Leu Ser Thr Leu Val Gly Arg Tyr Gly Arg Arg Thr Lys Leu Tyr
1310 1315 1320

Asn Ala Ser His Ser Asp Val Arg Asp Ser Leu Ala Arg Phe Ile
1325 1330 1335

Pro Ala Ile Gly Pro Val Gln Val Thr Thr Cys Glu Leu Tyr Glu
1340 1345 1350

Leu Glu Glu Ala Met Val Glu Lys Gly Gln Asp Gly Ser Ala Val
1355 1360 1365

Leu Glu Leu Asp Leu Cys Ser Arg Asp Val Ser Arg Ile Thr Phe
1370 1375 1380

Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr Ile Ala
1385 1390 1395

His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe
1400 1405 1410

Cys Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Ala Ile
1415 1420 1425

Leu Ala Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp
1430 1435 1440

Asp Thr Val Phe Ser Ala Ala Val Ala Ala Ala Lys Ala Ser Met
1445 1450 1455

Val Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn
1460 1465 1470

Phe Ser Leu Gly Leu Glu Cys Ala Ile Met Glu Glu Cys Gly Met
1475 1480 1485

Pro Gln Trp Leu Ile Arg Leu Tyr His Leu Ile Arg Ser Ala Trp
1490 1495 1500

Ile Leu Gln Ala Pro Lys Glu Ser Leu Arg Gly Phe Trp Lys Lys
1505 1510 1515

His Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr Val Trp Asn
1520 1525 1530

Met Ala Val Ile Thr His Cys Tyr Asp Phe Arg Asp Leu Gln Val
1535 1540 1545

Ala Ala Phe Lys Gly Asp Asp Ser Ile Val Leu Cys Ser Glu Tyr
1550 1555 1560

Arg Gln Ser Pro Gly Ala Ala Val Leu Ile Ala Gly Cys Gly Leu
1565 1570 1575

Lys Leu Lys Val Asp Phe Arg Pro Ile Gly Leu Tyr Ala Gly Val
1580 1585 1590

Val Val Ala Pro Gly Leu Gly Ala Leu Pro Asp Val Val Arg Phe
1595 1600 1605

Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro Gly Pro Glu Arg
1610 1615 1620

Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu Arg Lys Leu
1625 1630 1635

Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg Val Tyr
1640 1645 1650

Gly Val Ser Pro Gly Leu Val His Asn Leu Ile Glu Met Leu Gln
1655 1660 1665

Ala Val Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro
1670 1675 1680

Val Leu Asp Leu Thr Asn Ser Ile Leu Cys Arg Val Glu
1685 1690

(SEQ. ID NO.: 2)

Met Arg Pro Arg Pro Ile Leu Leu Leu Leu Leu Met Phe Leu Pro
1 5 10 15

Met Leu Pro Ala Pro Pro Pro Gly Gln Pro Ser Gly Arg Arg Arg
20 25 30

Gly Arg Arg Ser Gly Gly Ser Gly Gly Gly Phe Trp Gly Asp Arg
35 40 45

Val Asp Ser Gln Pro Phe Ala Ile Pro Tyr Ile His Pro Thr Asn
50 55 60

Pro Phe Ala Pro Asp Val Thr Ala Ala Ala Gly Ala Gly Pro Arg
65 70 75

Val Arg Gln Pro Ala Arg Pro Leu Gly Ser Ala Trp Arg Asp Gln
80 85 90

Ala Gln Arg Pro Ala Ala Ala Ser Arg Arg Arg Pro Thr Thr Ala
95 100 105

Gly Ala Ala Pro Leu Thr Ala Val Ala Pro Ala His Asp Thr Pro
110 115 120

Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg Gln
125 130 135

Tyr Asn Leu Ser Thr Ser Pro Leu Thr Ser Ser Val Ala Thr Gly
140 145 150

Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu Ser Pro Leu Leu Pro
155 160 165

Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala Thr Glu Ala Ser
170 175 180

Asn Tyr Ala Gln Tyr Arg Val Ala Arg Ala Thr Ile Arg Tyr Arg
185 190 195

```
Pro Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser Ile Ser
                200                 205                 210

Phe Typ Pro Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met Asn
                215                 220                 225

Ser Ile Thr Ser Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile
                230                 235                 240

Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn
                245                 250                 255

Gln Gly Trp Arg Ser Val Glu Thr Ser Gly Val Ala Glu Glu Glu
                260                 265                 270

Ala Thr Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Pro Val
                275                 280                 285

Asn Ser Tyr Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu
                290                 295                 300

Asp Phe Ala Leu Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn
                305                 310                 315

Thr Asn Thr Arg Val Ser Arg Tyr Ser Ser Thr Ala Arg His Arg
                320                 325                 330

Leu Arg Arg Gly Ala Asp Gly Thr Ala Glu Leu Thr Thr Thr Ala
                335                 340                 345

Ala Thr Arg Phe Met Lys Asp Leu Tyr Phe Thr Ser Thr Asn Gly
                350                 355                 360

Val Gly Glu Ile Gly Arg Gly Ile Ala Leu Thr Leu Phe Asn Leu
                365                 370                 375

Ala Asp Thr Leu Leu Gly Gly Leu Pro Thr Glu Leu Ile Ser Ser
                380                 385                 390

Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro Val Val Ser Ala Asn
                395                 400                 405

Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val Glu Asn Ala Gln
                410                 415                 420

Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp Leu Gly Glu
                425                 430                 435

Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln His Glu Gln Asp
                440                 445                 450

Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val Leu
                455                 460                 465

Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr
                470                 475                 480

Asp Gln Ser Thr Tyr Gly Ser Ser Thr Gly Pro Val Tyr Val Ser
                485                 490                 495

Asp Ser Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val
                500                 505                 510
```

```
Ala Arg Ser Leu Asp Trp Thr Lys Val Thr Leu Asp Gly Arg Pro
                515                 520                 525

Leu Ser Thr Ile Gln Gln Tyr Ser Lys Thr Phe Phe Val Leu Pro
            530                 535                 540

Leu Arg Gly Lys Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala
        545                 550                 555

Gly Tyr Pro Tyr Asn Tyr Asn Thr Thr Ala Ser Asp Gln Leu Leu
                560                 565                 570

Val Glu Asn Ala Ala Gly His Arg Val Ala Ile Ser Thr Tyr Thr
            575                 580                 585

Thr Ser Leu Gly Ala Gly Pro Val Ser Ile Ser Ala Val Ala Val
        590                 595                 600

Leu Ala Pro His Ser Val Leu Ala Leu Leu Glu Asp Thr Met Asp
    605                 610                 615

Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro Glu Cys
                620                 625                 630

Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val Ala
            635                 640                 645

Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Leu
        650                 655                 660
```

(SEQ. ID NO.: 3)

```
Met Asn Asn Met Ser Phe Ala Ala Pro Met Gly Ser Arg Pro Cys
1               5                   10                  15

Ala Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys
            20                  25                  30

Cys Pro Arg His Arg Pro Val Ser Arg Leu Ala Ala Val Val Gly
        35                  40                  45

Gly Ala Ala Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu
    50                  55                  60

Ile Leu Ser Pro Ser Gln Ser Pro Ile Phe Ile Gln Pro Thr Pro
65                  70                  75

Ser Pro Pro Met Ser Pro Leu Arg Pro Gly Leu Asp Leu Val Phe
            80                  85                  90

Ala Asn Pro Pro Asp His Ser Ala Pro Leu Gly Val Thr Arg Pro
        95                  100                 105

Ser Ala Pro Pro Leu Pro His Val Val Asp Leu Pro Gln Leu Gly
    110                 115                 120

Pro Arg Arg
```

[0022] The three-letter abbreviations follow the conventional amino acid shorthand for the twenty naturally occurring amino acids.

[0023] The recombinant HEV proteins may consist of at least one ORF protein but always include amine acids 1 to 660 of a hepatitis Evirus open reading frame 2 protein. Other recombinant proteins made up of more than one of the same or different ORF proteins may be made to alter the biological properties of the protein.

[0024] It is contemplated that additions, substitutions or deletions of discrete amino acids or of discrete sequences of amino acids may enhance the biological activity of the HEV proteins.

[0025] The present invention may use a nucleic acid sequence which is capable of directing the production of the

above-discussed HEV protein or proteins substantially homologous to the HEV proteins. This nucleic acid sequence, designated SAR-55, is set both below as SEQ ID NO.: 4 and was deposited with the American Type Culture Collection (ATCC) on September 17, 1992.

```
AGGCAGACCA CATATGTGGT CGATGCCATG GAGGCCCATC AGTTTATCAA    50
GGCTCCTGGC ATCACTACTG CTATTGAGCA GGCTGCTCTA GCAGCGGCCA   100
ACTCTGCCCT TGCGAATGCT GTGGTAGTTA GGCCTTTTCT CTCTCACCAG   150
CAGATTGAGA TCCTTATTAA CCTAATGCAA CCTCGCCAGC TTGTTTTCCG   200
CCCCGAGGTT TTCTGGAACC ATCCCATCCA GCGTGTTATC CATAATGAGC   250
TGGAGCTTTA CTGTCGCGCC CGCTCCGGCC GCTGCCTCGA AATTGGTGCC   300
CACCCCCGCT CAATAAATGA CAATCCTAAT GTGGTCCACC GTTGCTTCCT   350
CCGTCCTGCC GGGCGTGATG TTCAGCGTTG GTATACTGCC CCTACCCGCG   400
GGCCGGCTGC TAATTGCCGG CGTTCCGCGC TGCGCGGGCT CCCCGCTGCT   450
GACCGCACTT ACTGCTTCGA CGGGTTTTCT GGCTGTAACT TTCCCGCCGA   500
GACGGGCATC GCCCTCTATT CTCTCCATGA TATGTCACCA TCTGATGTCG   550
CCGAGGCTAT GTTCCGCCAT GGTATGACGC GGCTTTACGC TGCCCTCCAC   600
CTCCCGCCTG AGGTCCTGTT GCCCCCTGGC ACATACCGCA CCGCGTCGTA   650
CTTGCTGATC CATGACGGCA GGCGCGTTGT GGTGACGTAT GAGGGTGACA   700
CTAGTGCTGG TTATAACCAC GATGTTTCCA ACCTGCGCTC CTGGATTAGA   750
ACCACTAAGG TTACCGGAGA CCACCCTCTC GTCATCGAGC GGGTTAGGGC   800
CATTGGCTGC CACTTTGTCC TTTTACTCAC GGCTGCTCCG GAGCCATCAC   850
CTATGCCCTA TGTCCCTTAC CCCCGGTCTA CCGAGGTCTA TGTCCGATCG   900
ATCTTCGGCC CGGGTGGCAC CCCCTCCCTA TTTCCAACCT CATGCTCCAC   950
CAAGTCGACC TTCCATGCTG TCCCTGCCCA TATCTGGGAC CGTCTCATGT  1000
TGTTCGGGGC CACCCTAGAT GACCAAGCCT TTTGCTGCTC CGCCTAATG   1050
```

```
ACTTACCTCC GCGGCATTAG CTACAAGGTT ACTGTGGGCA CCCTTGTTGC 1100
CAATGAAGGC TGGAACGCCT CTGAGGACGC TCTTACAGCT GTCATCACTG 1150
CCGCCTACCT TACCATCTGC CACCAGCGGT ACCTCCGCAC TCAGGCTATA 1200
TCTAAGGGGA TGCGTCGCCT GGAGCGGGAG CATGCTCAGA AGTTTATAAC 1250
ACGCCTCTAC AGTTGGCTCT TTGAGAAGTC CGGCCGTGAT TATATCCCCG 1300
GCCGTCAGTT GGAGTTCTAC GCTCAGTGTA GGCGCTGGCT CTCGGCCGGC 1350
TTTCATCTTG ACCCACGGGT GTTGGTTTTT GATGAGTCGG CCCCCTGCCA 1400
CTGTAGGACT GCGATTCGTA AGGCGGTCTC AAAGTTTTGC TGCTTTATGA 1450
AGTGGCTGGG CCAGGAGTGC ACCTGTTTTC TACAACCTGC AGAAGGCGTC 1500
GTTGGCGACC AGGGCCATGA CAACGAGGCC TATGAGGGGT CTGATGTTGA 1550
CCCTGCTGAA TCCGCTATTA GTGACATATC TGGGTCCTAC GTAGTCCCTG 1600
GCACTGCCCT CCAACCGCTT TACCAAGCCC TTGACCTCCC CGCTGAGATT 1650
GTGGCTCGTG CAGGCCGGCT GACCGCCACA GTAAAGGTCT CCCAGGTCGA 1700
CGGGCGGATC GATTGTGAGA CCCTTCTCGG TAATAAAACC TTCCGCACGT 1750
CGTTTGTTGA CGGGGCGGTT TTAGAGACTA ATGGCCCAGA GCGCCACAAT 1800
CTCTCTTTTG ATGCCAGTCA GAGCACTATG GCCGCCGGCC CTTTCAGTCT 1850
CACCTATGCC GCCTCTGCTG CTGGGCTGGA GGTGCGCTAT GTCGCCGCCG 1900
GGCTTGACCA CCGGGCGGTT TTTGCCCCCG GCGTTTCACC CCGGTCAGCC 1950
CCTGGCGAGG TCACCGCCTT CTGTTCTGCC CTATACAGGT TTAATCGCGA 2000
GGCCCAGCGC CTTTCGCTGA CCGGTAATTT TTGGTTCCAT CCTGAGGGGC 2050
TCCTTGGCCC CTTTGCCCCG TTTTCCCCCG GGCATGTTTG GGAGTCGGCT 2100
AATCCATTCT GTGGCGAGAG CACACTTTAC ACCCGCACTT GGTCGGAGGT 2150
TGATGCTGTT CCTAGTCCAG CCCAGCCCGA CTTAGGTTTT ACATCTGAGC 2200
CTTCTATACC TAGTAGGGCC GCCACACCTA CCCCGGCGGC CCCTCTACCC 2250
CCCCCTGCAC CGGATCCTTC CCCTACTCTC TCTGCTCCGG CGCGTGGTGA 2300
GCCGGCTCCT GGCGCTACCG CCAGGGCCCC AGCCATAACC CACCAGACGG 2350
CCCGGCATCG CCGCCTGCTC TTTACCTACC CGGATGGCTC TAAGGTGTTC 2400
GCCGGCTCGC TGTTTGAGTC GACATGTACC TGGCTCGTTA ACGCGTCTAA 2450
TGTTGACCAC CGCCCTGGCG GTGGGCTCTG TCATGCATTT TACCAGAGGT 2500
ACCCCGCCTC CTTTGATGCT GCCTCTTTTG TGATGCGCGA CGGCGCGGCC 2550
GCCTACACAT TAACCCCCCG GCCAATAATT CATGCCGTCG CTCCTGATTA 2600
TAGGTTGGAA CATAACCCAA AGAGGCTTGA GGCTGCCTAC CGGGAGACTT 2650
GCTCCCGCCT CGGTACCGCT GCATACCCAC TCCTCGGGAC CGGCATATAC 2700
CAGGTGCCGA TCGGTCCCAG TTTTGACGCC TGGGAGCGGA ATCACCGCCC 2750
CGGGGACGAG TTGTACCTTC CTGAGCTTGC TGCCAGATGG TTCGAGGCCA 2800
ATAGGCCGAC CTGCCCAACT CTCACTATAA CTGAGGATGT TGCGCGGACA 2850
```

GCAAATCTGG CTATCGAACT TGACTCAGCC ACAGACGTCG GCCGGGCCTG 2900
TGCCGGCTGT CGAGTCACCC CCGGCGTTGT GCAGTACCAG TTTACCGCAG 2950
GTGTGCCTGG ATCCGGCAAG TCCCGCTCTA TTACCCAAGC CGACGTGGAC 3000
GTTGTCGTGG TCCCGACCCG GGAGTTGCGT AATGCCTGGC GCCGCCGCGG 3050
CTTCGCTGCT TTCACCCCGC ACACTGCGGC TAGAGTCACC CAGGGGCGCC 3100
GGGTTGTCAT TGATGAGGCC CCGTCCCTTC CCCCTCATTT GCTGCTGCTC 3150
CACATGCAGC GGGCCGCCAC CGTCCACCTT CTTGGCGACC CGAATCAGAT 3200
CCCAGCCATC GATTTTGAGC ACGCCGGGCT CGTTCCCGCC ATCAGGCCCG 3250
ATTTGGCCCC CACCTCCTGG TGGCATGTTA CCCATCGCTG CCCTGCGGAT 3300
GTATGTGAGC TAATCCGCGG CGCATACCCT ATGATTCAGA CCACTAGTCG 3350
GGTCCTCCGG TCGTTGTTCT GGGGTGAGCC CGCCGTTGGG CAGAAGCTAG 3400
TGTTCACCCA GGCGGCTAAG GCCGCCAACC CCGGTTCAGT GACGGTCCAT 3450
GAGGCACAGG GCGCTACCTA CACAGAGACT ACCATCATTG CCACGGCAGA 3500
TGCTCGAGGC CTCATTCAGT CGTCCCGAGC TCATGCCATT GTTGCCTTGA 3550
CGCGCCACAC TGAGAAGTGC GTCATCATTG ACGCACCAGG CCTGCTTCGC 3600
GAGGTGGGCA TCTCCGATGC AATCGTTAAT AACTTTTTCC TTGCTGGTGG 3650
CGAAATTGGC CACCAGCGCC CATCTGTTAT CCCTCGCGGC AATCCTGACG 3700
CCAATGTTGA CACCTTGGCT GCCTTCCCGC CGTCTTGCCA GATTAGCGCC 3750
TTCCATCAGT TGGCTGAGGA GCTTGGCCAC AGACCTGCCC CTGTCGCGGC 3800
TGTTCTACCG CCCTGCCCTG AGCTTGAACA GGGCCTTCTC TACCTGCCCC 3850
AAGAACTCAC CACCTGTGAT AGTGTCGTAA CATTTGAATT AACAGATATT 3900
GTGCATTGTC GTATGGCCGC CCCGAGCCAG CGCAAGGCCG TGCTGTCCAC 3950
GCTCGTGGGC CGTTATGGCC GCCGCACAAA GCTCTACAAT GCCTCCCACT 4000
CTGATGTTCG CGACTCTCTC GCCCGTTTTA TCCCGGCCAT TGGCCCCGTA 4050
CAGGTTACAA CCTGTGAATT GTACGAGCTA GTGGAGGCCA TGGTCGAGAA 4100
GGGCCAGGAC GGCTCCGCCG TCCTTGAGCT CGACCTTTGT AGCCGCGACG 4150
TGTCCAGGAT CACCTTCTTC CAGAAAGATT GTAATAAATT CACCACGGGG 4200
GAGACCATCG CCCATGGTAA AGTGGGCCAG GGCATTTCGG CCTGGAGTAA 4250
GACCTTCTGT GCCCTTTTCG GCCCCTGGTT CCGTGCTATT GAGAAGGCTA 4300
TCCTGGCCCT GCTCCCTCAG GGTGTGTTTT ATGGGGATGC CTTTGATGAC 4350
ACCGTCTTCT CGGCGGCTGT GGCCGCAGCA AAGGCATCCA TGGTGTTCGA 4400
GAATGACTTT TCTGAGTTTG ATTCCACCCA GAATAATTTT TCCTTGGGCC 4450
TAGAGTGTGC TATTATGGAG GAGTGTGGGA TGCCGCAGTG GCTCATCCGC 4500
TTGTACCACC TTATAAGGTC TGCGTGGATT CTGCAGGCCC CGAAGGAGTC 4550
CCTGCGAGGG TTTTGGAAGA AACACTCCGG TGAGCCCGGC ACCCTTCTGT 4600
GGAATACTGT CTGGAACATG GCCGTTATCA CCCACTGTTA TGATTTCCGC 4650

```
GATCTGCAGG TGGCTGCCTT TAAAGGTGAT GATTCGATAG TGCTTTGCAG 4700
TGAGTACCGT CAGAGCCCAG GGGCTGCTGT CCTGATTGCT GGCTGTGGCC 4750
TAAAGTTGAA GGTGGATTTC CGTCCGATTG GTCTGTATGC AGGTGTTGTG 4800
GTGGCCCCCG GCCTTGGCGC GCTTCCTGAT GTCGTGCGCT TCGCCGGTCG 4850
GCTTACTGAG AAGAATTGGG GCCCTGGCCC CGAGCGGGCG GAGCAGCTCC 4900
GCCTCGCTGT GAGTGATTTT CTCCGCAAGC TCACGAATGT AGCTCAGATG 4950
TGTGTGGATG TTGTCTCTCG TGTTTATGGG GTTTCCCCTG GGCTCGTTCA 5000
TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG GCTCATTTCA 5050
CTGAGTCAGT GAAGCCAGTG CTTGACCTGA CAAATTCAAT TCTGTGTCGG 5100
GTGGAATGAA TAACATGTCT TTTGCTGCGC CCATGGGTTC GCGACCATGC 5150
GCCCTCGGCC TATTTTGCTG TTGCTCCTCA TGTTTCTGCC TATGCTGCCC 5200
GCGCCACCGC CCGGTCAGCC GTCTGGCCGC CGTCGTGGGC GGCGCAGCGG 5250
CGGTTCCGGC GGTGGTTTCT GGGGTGACCG GGTTGATTCT CAGCCCTTCG 5300
CAATCCCCTA TATTCATCCA ACCAACCCCT TCGCCCCCGA TGTCACCGCT 5350
GCGGCCGGGG CTGGACCTCG TGTTCGCCAA CCCGCCCGAC CACTCGGCTC 5400
CGCTTGGCGT GACCAGGCCC AGCGCCCCGC CGCTGCCTCA CGTCGTAGAC 5450
CTACCACAGC TGGGGCCGCG CCGCTAACCG CGGTCGCTCC GGCCCATGAC 5500
ACCCCGCCAG TGCCTGATGT TGACTCCCGC GGCGCCATCC TGCGCCGGCA 5550
GTATAACCTA TCAACATCTC CCCTCACCTC TTCCGTGGCC ACCGGCACAA 5600
ATTTGGTTCT TTACGCCGCT CCTCTTAGCC CGCTTCTACC CCTCCAGGAC 5650
GGCACCAATA CTCATATAAT GGCTACAGAA GCTTCTAATT ATGCCCAGTA 5700
CCGGGTTGCT CGTGCCACAA TTCGCTACCG CCCGCTGGTC CCCAACGCTG 5750
TTGGTGGCTA CGCTATCTCC ATTTCGTTCT GGCCACAGAC CACCACCACC 5800
CCGACGTCCG TTGACATGAA TTCAATAACC TCGACGGATG TCCGTATTTT 5850
AGTCCAGCCC GGCATAGCCT CCGAGCTTGT TATTCCAAGT GAGCGCCTAC 5900
ACTATCGCAA CCAAGGTTGG CGCTCTGTTG AGACCTCCGG GGTGGCGGAG 5950
GAGGAGGCCA CCTCTGGTCT TGTCATGCTC TGCATACATG GCTCACCTGT 6000
AAATTCTTAT ACTAATACAC CCTATACCGG TGCCCTCGGG CTGTTGGACT 6050
TTGCCCTCGA ACTTGAGTTC CGCAACCTCA CCCCCGGTAA TACCAATACG 6100
CGGGTCTCGC GTTACTCCAG CACTGCCCGT CACCGCCTTC GTCGCGGTGC 6150
AGATGGGACT GCCGAGCTCA CCACCACGGC TGCTACTCGC TTCATGAAGG 6200
ACCTCTATTT TACTAGTACT AATGGTGTTG GTGAGATCGG CCGCGGGATA 6250
GCGCTTACCC TGTTTAACCT TGCTGACACC CTGCTTGGCG GTCTACCGAC 6300
AGAATTGATT TCGTCGGCTG GTGGCCAGCT GTTCTACTCT CGCCCCGTCG 6350
TCTCAGCCAA TGGCGAGCCG ACTGTTAAGC TGTATACATC TGTGGAGAAT 6400
GCTCAGCAGG ATAAGGGTAT TGCAATCCCG CATGACATCG ACCTCGGGGA 6450
```

```
ATCCCGTGTA GTTATTCAGG ATTATGACAA CCAACATGAG CAGGACCGAC  6500
CGACACCTTC CCCAGCCCCA TCGCGTCCTT TTTCTGTCCT CCGAGCTAAC  6550
GATGTGCTTT GGCTTTCTCT CACCGCTGCC GAGTATGACC AGTCCACTTA  6600
CGGCTCTTCG ACCGGCCCAG TCTATGTCTC TGACTCTGTG ACCTTGGTTA  6650
ATGTTGCGAC CGGCGCGCAG GCCGTTGCCC GGTCACTCGA CTGGACCAAG  6700
GTCACACTTG ATGGTCGCCC CCTTTCCACC ATCCAGCAGT ATTCAAAGAC  6750
CTTCTTTGTC CTGCCGCTCC GCGGTAAGCT CTCCTTTTGG GAGGCAGGAA  6800
CTACTAAAGC CGGGTACCCT TATAATTATA ACACCACTGC TAGTGACCAA  6850
CTGCTCGTTG AGAATGCCGC TGGGCATCGG GTTGCTATTT CCACCTACAC  6900
TACTAGCCTG GGTGCTGGCC CCGTCTCTAT TTCCGCGGTT GCTGTTTTAG  6950
CCCCCCACTC TGTGCTAGCA TTGCTTGAGG ATACCATGGA CTACCCTGCC  7000
CGCGCCCATA CTTTCGATGA CTTCTGCCCG GAGTGCCGCC CCCTTGGCCT  7050
CCAGGGTTGT GCTTTTCAGT CTACTGTCGC TGAGCTTCAG CGCCTTAAGA  7100
TGAAGGTGGG TAAAACTCGG GAGTTATAGT TTATTTGCTT GTGCCCCCCT  7150
TCTTTCTGTT GCTTATTT                                     7168
```

[0026] The abbreviations used for the nucleotides are those standardly used in the art.

[0027] In a preferred embodiment, the nucleic acid sequence used in the method of the present invention consists of nucleotides 5147 to 7126 of SEQ ID NO:4.

[0028] The sequence in one direction has been designated by convention as the "plus" sequence since it is the protein-encoding strand of RNA viruses and this is the sequence shown above as SEQ ID. NO.:4.

[0029] The deduced amino acid sequences of the open reading frames of SAR-55 have SEQ ID NO. 1, SEQ ID NO. 2, and SEQ ID NO. 3. ORF-1 starts at nucleotide 28 of SEQ. ID NO. 4 and extends 5078 nucleotides; ORF-2 starts at nucleotide 5147 of SEQ. ID NO. 4 and extends 1979 nucleotides; and ORF-3 starts at nucleotide 5106 of SEQ. ID NO. 4 and extends 368 nucleotides.

[0030] Variations are contemplated in the DNA sequence which will result in a DNA sequence that is capable of directing production of analogs of the ORF-2 protein. It should be noted that the DNA sequence set forth above may represent an embodiment of the present invention. Due to the degeneracy of the genetic code, it is to be understood that numerous choices of nucleotides may be made that will lead to a DNA sequence capable of directing production of the ORF proteins or their analogs, including ORF2 proteins. As such, DNA sequences which are functionally equivalent to the sequences set forth above or which are functionally equivalent to sequences that would direct production of analogs of the ORF proteins produced pursuant to the amino acid sequence set forth above, are intended to be encompassed within the present invention.

[0031] The present invention discloses a method for detecting the hepatitis E virus in biological samples based on selective amplification of hepatitis E gene fragments. Preferably, this method utilizes a pair of single-stranded primers derived from non-homologous regions of opposite strands of a DNA duplex fragment, which in turn is derived from a hepatitis E virus whose genome contains a region homologous to the SAR-55 sequence shown in SEQ ID No.: 4. These primers can be used in a method following the process for amplifying selected nucleic acid sequences as defined in U.S. Patent No. 4,683,202.

[0032] The present invention also discloses the use of single-stranded antisense poly- or oligonucleotides derived from sequences homologous to the SAR-55 cDNA to inhibit the expression of hepatitis E genes. These anti-sense poly- or oligonucleotides can be either DNA or RNA. The targeted sequence is typically messenger RNA and more preferably, a signal sequence required for processing or translation of the RNA. The antisense poly- or oligonucleotides can be conjugated to a polycation such as polylysine as disclosed in Lemaitre, M. et al. (1989) Proc Natl Acad Sci USA 84: 648-652; and this conjugate can be administered to a mammal in an amount sufficient to hybridize to and inhibit the function of the messenger RNA.

[0033] The present invention includes a recombinant DNA method for the manufacture of HEV proteins, preferably a protein composed of at least one ORF protein, including the ORF-2 protein most preferably at least one ORF-2 protein. The recombinant ORF protein may be composed of one ORF protein or a combination of the same or different ORF

proteins, including the ORF-2 protein. A natural or synthetic nucleic acid sequence may be used to direct production of the HEV proteins.

[0034] In one embodiment, such a method comprises:

(a) preparation of a nucleic acid sequence capable of directing a host organism to produce a protein of HEV; including the ORF-2 protein
(b) cloning the nucleic acid sequence into a vector capable of being transferred into and replicated in a host organism, such vector containing operational elements for the nucleic acid sequence;
(c) transferring the vector containing the nucleic acid and operational elements into a host organism capable of expressing the protein;
(d) culturing the host organism under conditions appropriate for amplification of the vector and expression of the protein; and
(e) harvesting the protein.

[0035] In another embodiment of the invention the method for the recombinant DNA synthesis of a protein including the ORF-2 protein encoded by nucleic acids of HEV, preferably encoding at least one ORF of HEV or a combination of the same or different ORF proteins including the ORF-2 protein, most preferably encoding at least one ORF-2 protein, comprises:

(a) culturing a transformed or transfected host organism containing a nucleic acid sequence capable of directing the host organism to produce a protein, under condition such that the protein is produced, said protein exhibiting substantial homology to a native HEV protein isolated from HEV having the amino acid sequence according to SEQ ID NO. 1, SEQ ID NO. 2 or SEQ ID NO. 3, or combinations thereof, however, always including the amino acid sequence according to SEQ ID NO. 2.

[0036] In one embodiment, the RNA sequence of the viral genome of HEV strain SAR-55 was isolated and cloned to cDNA as follows. Viral RNA is extracted from a biological sample collected from cynomolgus monkeys infected with SAR-55 and the viral RNA is then reverse transcribed and amplified by polymerase chain reaction using primers complementary to the plus or minus strands of the genome of a strain of HEV from Burma (Tam et al.) or the SAR-55 genome The PCR fragments are subcloned into pBR322 or pGEM-3 and the double-stranded PCR fragments were sequenced.

[0037] The vectors contemplated for use in the present invention include a baculo virus expression vector into which a nucleic acid sequence as described above can be inserted, along with any preferred or required operational elements, and which vector can then be subsequently transferred into a host organism and replicated in such organism. The vector used is one whose restriction sites have been well documented and which contains the operational elements preferred or required for transcription of the nucleic acid sequence.

[0038] The "operational elements" as discussed herein include at least one promoter, at least one operator, at least one leader sequence, at least one terminator codon, and any other DNA sequences necessary or preferred for appropriate transcription and subsequent translation of the vector nucleic acid. In particular, it is contemplated that such vectors will contain at least one origin of replication recognized by the host organism along with at least one selectable marker and at least one promoter sequence capable of initiating transcription of the nucleic acid sequence.

[0039] In construction of the cloning vector of the present invention, it should additionally be noted that multiple copies of the nucleic acid sequence and its attendant operational elements may be inserted into each vector. In such an embodiment, the host organism would produce greater amounts per vector of the desired HEV protein. The number of multiple copies of the DNA sequence which may be inserted into the vector is limited only by the ability of the resultant vector due to its size, to be transferred into and replicated and transcribed in an appropriate host microorganism.

[0040] In another embodiment, restriction digest fragments containing a coding sequence for HEV proteins can be inserted into a suitable expression vector that functions in prokaryotic or eukaryotic cells. By suitable is meant that the vector is capable of carrying and expressing a complete nucleic acid sequence coding for HEV proteins, preferably at least one complete ORF protein. In the case of ORF-2, the expressed protein should form viral-like particles. The expression vector is a baculovirus expression vector that functions in a eukaryotic cell, namely an insect cell. Referred is the baculovirus transfer vector, pBlueBac. Preferred vectors are p63-2, which contains the complete ORF-2 gene, and P59-4, which contains the complete ORF-3 and ORF-2 genes. These vectors were deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852 USA on September 10, 1992. Example 1 illustrates the cloning of the ORF-2 gene into pBluebac to produce p63-2. This method includes digesting the genome of HEV strain SAR-55 with the restriction enzymes NruI and BglII, inserting a polylinker containing BlnI and BglII sites into the unique NheI site of the vector and inserting the NruI-BglII ORF-2 fragment in BlnI-BglII pBluebac using an adapter.

[0041] In yet another embodiment, the selected recombinant expression vector may then be transfected into a suitable eukaryotic cell system for purposes of expressing the recombinant protein. A preferred eukaryotic cell system is SF9

insect cells. One preferred method involves use of the Pbluebac expression vector where the insect cell line SF9 is cotransfected with recombinant pBluebac and AcMNPV baculovirus DNA by the Ca precipitation method.

[0042] The expressed recombinant protein may be detected by methods known in the art which include Coomassie blue staining and Western blotting using sera containing anti-HEV antibody as shown in Example 2. Another method is the detection of virus-like particles by immunoelectron microscopy as shown in Example 3.

[0043] The recombinant protein expressed by the SF9 cells can be obtained as a crude lysate or it can be purified by standard protein purification procedures known in the art which may include differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity, and immunoaffinity chromatography and the like. In the case of immunoaffinity chromatography, the recombinant protein may be purified by passage through a column containing a resin which has bound thereto antibodies specific for the ORF protein.

[0044] The expressed recombinant proteins of this invention can be used in immunoassays for diagnosing or prognosing hepatitis E in a mammal including but not limited to humans, chimpanzees, Old World monkeys, New World monkeys, other primates and the like. In a preferred embodiment, the immunoassay is useful in diagnosing hepatitis E infection in humans. Immunoassays using the HEV proteins, particularly the ORF proteins, including the ORF2 protein and especially ORF 2 proteins, provide a highly specific, sensitive and reproducible method for diagnosing HEV infections, in contrast to immunoassays which utilize partial ORF proteins.

[0045] Immunoassays mentioned in the present invention may be a radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme immunoassay, chemiluminescent assay, immunohistochemical assay and the like. Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Edition, Rose and Bigazzi, eds., John Wiley and Sons, 1980 and Campbell et al., Methods of Immunology, W.A. Benjamin, Inc., 1964, both of which are incorporated herein by reference. Such assays may be a direct, indirect, competitive, or noncompetitive immunoassay as described in the art. (Oellerich, M. 1984. J.Clin. Chem. Clin. BioChem. 22: 895-904) Biological samples appropriate for such detection assays include, but are not limited to, tissue biopsy extracts, whole blood, plasma, serum, cerebrospinal fluid, pleural fluid, urine and the like.

[0046] The test serum is reacted with a solid phase reagent having surface-bound recombinant HEV protein as an antigen, preferably the ORF-2 protein or combination of different ORF proteins such as ORF-2 and ORF-3, however always including ORF-2. More preferably, the HEV protein is an ORF-2 protein that forms virus-like particles. The solid surface reagent can be prepared by known techniques for attaching protein to solid support material. These attachment methods include non-specific adsorption of the protein to the support or covalent attachment of the protein to a reactive group on the support. After reaction of the antigen with anti-HEV antibody, unbound serum components are removed by washing and the antigen-antibody complex is reacted with a secondary antibody such as labelled anti-human antibody. The label may be an enzyme which is detected by incubating the solid support in the presence of a suitable fluorimetric or colorimetric reagent. Other detectable labels may also be used, such as radiolabels or colloidal gold, and the like.

[0047] In a preferred embodiment, protein expressed by the recombinant vector pBluebac containing the entire ORF-2 sequence of SAR-55 is used as a specific binding agent to detect anti-HEV antibodies, preferably IgG or IgM antibodies. Examples 4 and 5 show the results of an ELISA in which the solid phase reagent has recombinant ORF-2 as the surface antigen. This protein, encoded by the entire ORF-2 nucleic acid sequence, is superior to the partial ORF-2 proteins, as it is reactive with more antisera from different primate species infected with HEV than are partial antigens of ORF-2. The protein of the present invention is also capable of detecting antibodies produced in response to different strains of HEV but not to Hepatitis A, B, C or D viruses.

[0048] The HEV protein and analogs may be prepared in the form of a kit, alone, or in combinations with other reagents such as secondary antibodies, for use in immunoassays.

[0049] The recombinant HEV proteins, including an ORF-2 protein, preferably an ORF-2 protein or combination of ORF proteins, and substantially homologous proteins and analogs of the invention can be used as a vaccine to protect mammals against challenge with Hepatitis Evirus. The vaccine, which acts as an immunogen, may be a cell, cell lysate from cells transfected with a recombinant expression vector or a culture supernatant containing the expressed protein. Alternatively, the immunogen is a partially or substantially purified recombinant protein. While it is possible for the immunogen to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation.

[0050] The formulations disclosed in the present invention, both for veterinary and for human use, comprise an immunogen as described above, together with one or more pharmaceutically acceptable carriers and optionally other therapeutic ingredients. The carrier(s) must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The formulations may conveniently be presented in unit dosage form and may be prepared by any method well-known in the pharmaceutical art.

[0051] All methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired formulation.

[0052] Formulations suitable for intravenous, intramuscular, subcutaneous, or intraperitoneal administration conveniently comprise sterile aqueous solutions of the active ingredient with solutions which are preferably isotonic with the blood of the recipient. Such formulations may be conveniently prepared by dissolving solid active ingredient in water containing physiologically compatible substances such as sodium chloride (e.g. 0.1-2.OM), glycine, and the like, and having a buffered pH compatible with physiological conditions to produce an aqueous solution, and rendering said solution sterile. These may be present in unit or multi-dose containers, for example, sealed ampoules or vials.

[0053] The formulations disclosed in the present invention may incorporate a stabilizer. Illustrative stabilizers are polyethylene glycol, proteins, saccharides, amino acids, inorganic acids, and organic acids which may be used either on their own or as admixtures. These stabilizers are preferably incorporated in an amount of 0.11-10,000 parts by weight per part by weight of immunogen. If two or more stabilizers are to be used, their total amount is preferably within the range specified above. These stabilizers are used in aqueous solutions at the appropriate concentration and pH. The specific osmotic pressure of such aqueous solutions is generally in the range of 0.1-3.0 osmoles, preferably in the range of 0.8-1.2. The pH of the aqueous solution is adjusted to be within the range of 5.0-9.0, preferably within the range of 6-8. In formulating the immunogen disclosed in the present invention, anti-adsorption agent may be used.

[0054] Additional pharmaceutical methods may be employed to control the duration of action. Controlled release preparations may be achieved through the use of polymer to complex or absorb the proteins or their derivatives. The controlled delivery may be exercised by selecting appropriate macromolecules (for example polyester, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine sulfate) and the concentration of macromolecules as well as the methods of incorporation in order to control release. Another possible method to control the duration of action by controlled-release preparations is to incorporate the proteins, protein analogs or their functional derivatives, into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly (lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxy-methylcellulose or gelatin-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

[0055] When oral preparations are desired, the compositions may be combined with typical carriers, such as lactose, sucrose, starch, talc magnesium stearate, crystalline cellulose, methyl cellulose, carboxymethyl cellulose, glycerin, sodium alginate or gum arabic among others.

[0056] The proteins as a result of the method of the present invention may be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition as described above.

[0057] Vaccination can be conducted by conventional methods. For example, the immunogen can be used in a suitable diluent such as saline or water, or complete or incomplete adjuvants. Further, the immunogen may or may not be bound to a carrier to make the protein immunogenic. Examples of such carrier molecules include but are not limited to bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), tetanus toxoid, and the like. The immunogen can be administered by any route appropriate for antibody production such as intravenous, intraperitoneal, intramuscular, subcutaneous, and the like. The immunogen may be administered once or at periodic intervals until a significant titer of anti-HEV antibody is produced. The antibody may be detected in the serum using an immunoassay.

[0058] The administration of the immunogen as a result of the method of the present invention may be for either a prophylactic or therapeutic purpose. When provided prophylactically, the immunogen is provided in advance of any exposure to HEV or in advance of any symptom due to HEV infection. The prophylactic administration of the immunogen serves to prevent or attenuate any subsequent infection of HEV in a mammal. When provided therapeutically, the immunogen is provided at (or shortly after) the onset of the infection or at the onset of any symptom of infection or disease caused by HEV. The therapeutic administration of the immunogen serves to attenuate the infection or disease.

[0059] Preferably mentioned is a vaccine prepared using recombinant ORF-2 protein expressed by the ORF-2 sequence of HEV strain SAR-55 and equivalents thereof. Since the recombinant ORF-2 protein has already been demonstrated to be reactive with a variety of HEV-positive sera, their utility in protecting against a variety of HEV strains is indicated.

[0060] In addition to use as a vaccine, the compositions can be used to prepare antibodies to HEV virus-like particles. The antibodies can be used directly as antiviral agents. To prepare antibodies, a host animal is immunized using the virus particles or, as appropriate, non-particle antigens native to the virus particle are bound to a carrier as described above for vaccines. The host serum or plasma is collected following an appropriate time interval to provide a composition comprising antibodies reactive with the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

[0061] The antibody compositions can be made even more compatible with the host system by minimizing potential adverse immune system responses. This is accomplished by removing all or a portion of the Fc portion of a foreign

species antibody or using an antibody of the same species as the host animal, for example, the use of antibodies from human/human hybridomas. Humanized antibodies (i.e., nonimmunogenic in a human) may be produced, for example, by replacing an immunogenic portion of an antibody with a corresponding, but nonimmunogenic portion (i.e., chimeric antibodies). Such chimeric antibodies may contain the reactive or antigen binding portion of an antibody from one species and the Fc portion of an antibody (nonimmunogenic) from a different species. Examples of chimeric antibodies, include but are not limited to, non-human mammal-human chimeras, rodent-human chimeras, murine-human and rat-human chimeras (Robinson et al., International Patent Application 184,187; Taniguchi M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., PCT Application WO 86/01533; Cabilly et al., 1987 Proc. Natl. Acad. Sci. USA 84:3439; Nishimura et al., 1987 Canc. Res. 47:999; Wood et al., 1985 Nature 314:446; Shaw et al., 1988 J. Natl. Cancer Inst. 80: 15553, all incorporated herein by reference).

[0062] General reviews of "humanized" chimeric antibodies are provided by Morrison S., 1985 Science 229:1202 and by oi et al., 1986 BioTechniques 4:214.

[0063] Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones et al., 1986 Nature 321:552; Verhoeyan et al., 1988 Science 239:1534; Biedleret al. 1988 J. Immunol. 141:4053, all incorporated herein by reference).

[0064] The antibodies or antigen binding fragments may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of the PCT patent applications; publication number WO 901443, WO901443, and WO 9014424 and in Huse et al., 1989 Science 246:1275-1281.

[0065] The antibodies can also be used as a means of enhancing the immune response. The antibodies can be administered in amounts similar to those used for other therapeutic administrations of antibody. For example, pooled gamma globulin is administered at 0.02-0.1 ml/lb body weight during the early incubation period of other viral diseases such as rabies, measles and hepatitis B to interfere with viral entry into cells. Thus, antibodies reactive with the HEV virus particle can be passively administered alone or in conjunction with another anti-viral agent to a host infected with an HEV to enhance the immune response and/or the effectiveness of an antiviral drug.

[0066] Alternatively, anti-HEV antibodies can be induced by administering anti-idiotype antibodies as immunogens. Conveniently, a purified anti-HEV antibody preparation prepared as described above is used to induce anti-idiotype antibody in a host animal. The composition is administered to the host animal in a suitable diluent. Following administration, usually repeated administration, the host produces anti-idiotype antibody. To eliminate an immunogenic response to the Fc region, antibodies produced by the same species as the host animal can be used or the FC region of the administered antibodies can be removed. Following induction of anti-idiotype antibody in the host animal, serum or plasma is removed to provide an antibody composition. The composition can be purified as described above for anti-HEV antibodies, or by affinity chromatography using anti-HEV antibodies bound to the affinity matrix. The anti-idiotype antibodies produced are similar in conformation to the authentic HEV-antigen and may be used to prepare an HEV vaccine rather than using an HEV particle antigen.

[0067] When used as a means of inducing anti-HEV virus antibodies in an animal, the manner of injecting the antibody is the same as for vaccination purposes, namely intramuscularly, intraperitoneally, subcutaneously or the like in an effective concentration in a physiologically suitable diluent with or without adjuvant. One or more booster injections may be desirable.

[0068] The HEV derived proteins as a result of the method of the invention are also intended for use in producing antiserum designed for pre-or post-exposure prophylaxis. Here an HEV protein, or mixture of proteins is formulated with a suitable adjuvant and administered by injection to human volunteers, according to known methods for producing human antisera. Antibody response to the injected proteins is monitored, during a several-week period following immunization, by periodic serum sampling to detect the presence of anti-HEV serum antibodies, using an immunoassay as described herein.

[0069] The antiserum from immunized individuals may be administered as a pre-exposure prophylactic measure for individuals who are at risk of contracting infection. The antiserum is also useful in treating an individual post-exposure, analogous to the use of high titer antiserum against hepatitis B virus for post-exposure prophylaxis.

[0070] For both in vivo use of antibodies to HEV virus-like particles and proteins and anti-idiotype antibodies and diagnostic use, it may be preferable to use monoclonal antibodies. Monoclonal anti-virus particle antibodies or anti-idiotype antibodies can be produced as follows. The spleen or lymphocytes from an immunized animal are removed and immortalized or used to prepare hybridomas by methods known to those skilled in the art. (Goding, J.W. 1983. Monoclonal Antibodies: Principles and Practice, Pladermic Press, Inc., NY, NY, pp. 56-97). To produce a human-human hybridoma, a human lymphocyte donor is selected. A donor known to be infected with HEV (where infection has been shown for example by the presence of anti-virus antibodies in the blood or by virus culture) may serve as a suitable lymphocyte donor. Lymphocytes can be isolated from a peripheral blood sample or spleen cells may be used if the donor is subject to splenectomy. Epstein-Barr virus (EBV) can be used to immortalize human lymphocytes or a human fusion partner can be used to produce human-human hybridomas. Primary in vitro immunization with peptides can also be used in the generation of human monoclonal antibodies.

[0071] Antibodies secreted by the immortalized cells are screened to determine the clones that secrete antibodies of the desired specificity. For monoclonal anti-virus particle antibodies, the antibodies must bind to HEV virus particles. For monoclonal anti-idiotype antibodies, the antibodies must bind to anti-virus particle antibodies. Cells producing antibodies of the desired specificity are selected.

[0072] The above described antibodies and antigen binding fragments thereof may be supplied in kit form alone, or as a pharmaceutical composition for in vivo use. The antibodies may be used for therapeutic uses, diagnostic use in immunoassays or as an immunoaffinity agent to purify ORF proteins as described herein.

[0073] The materials used in the following Examples were as follows:

Primates. Chimpanzee (Chimp) (*Pan troglodytes*). Old world monkeys: cynomolgus monkeys (Cyno) (*Macaca fascicularis*), rhesus monkeys (Rhesus) (*M. mulatta*), pigtail monkeys (PT) (M. nemestrina), and African green monkeys (AGM) (*Cercopithecus aethiops*). New World monkeys: mustached tamarins (Tam) (*Saguinus mystax*), squirrel monkeys (SQM) (Saimiri sciureus) and owl monkeys (OWL) (*Aotus trivigatus*). Primates were housed singly under conditions of biohazard containment. The housing, maintenance and care of the animals met or exceeded all requirements for primate husbandry.

[0074] Most animals were inoculated intravenously with HEV, strain SAR-55 contained in 0.5 ml of stool suspension diluted in fetal calf serum as described in Tsarev, S.A. et al. (1992), Proc. Natl. Acad. Sci USA, 89:559-563; and Tsarev, S.A. et al. (1992), J. Infect. Dis. (submitted). Chimp-1313 and 1310 were inoculated with a pool of stools collected from 7 Pakistani hepatitis E patients.

[0075] Serum samples were collected approximately twice a week before and after inoculation. Levels of the liver enzymes serum alanine amino transferase (ALT), isocitrate dehydrogenase (ICD), and gamma glutamyl transferase (GGT) were assayed with commercially available tests (Medpath Inc., Rockville, MD). Serologic tests were performed as described above.

EXAMPLE 1

Identification of the DNA Sequence of the Genome of HEV Strain SAR-55.

[0076] *Preparation of Virus RNA Template for PCR.* Bile from an HEV-infected cynomolgus monkey (10 μl), 20% (wt/vol) SDS (to a final concentration of 1%), proteinase K (10 mg/ml; to a final concentration of 1 mg/ml), 1 μl of tRNA (10 mg/ml), and 3 μl of 0.5 M EDTA were mixed in a final volume of 250 μl and incubated for 30 min. at 55˚C. Total nucleic acids were extracted from bile twice with phenol/chloroform, 1:1 (vol/vol), at 65˚C and once with chloroform, then precipitated by ethanol, washed with 95% ethanol, and used for RT-PCR. RT-PCR amplification of HEV RNA from feces and especially from sera was more efficient when RNA was more extensively purified. Serum (100 μl) or a 10% fecal suspension (200 μl) was treated as above with proteinase K. After a 30-min incubation, 300 μl of CHAOS buffer (4.2 M guanidine thiocyanate/0.5 *N*-lauroylsarocosine/0.025 M Tris-HCl, pH 8.0) was added. Nucleic acids were extracted twice with phenol/chloroform at 65˚C followed by chloroform extraction at room temperature. Then 7.5 M ammonium acetate (225 μl) was added to the upper phase and nucleic acids were precipitated with 0.68 ml of 2-propanol. The pellet was dissolved in 300 ul CHAOS buffer and 100 ul of H_2O was added. Chloroform extraction and 2-propanol precipitation were repeated. Nucleic acids were dissolved in water, precipitated with ethanol, washed with 95% ethanol, and used for RT-PCR.

[0077] *Primers.* Ninety-four primers, 21-40 nucleotides (nt) long, and complementary to plus or minus strands of the genome of a strain of HEV from Burma (BUR-121) (Tam, A.W. et al. (1991), Virology, 185:120-131) or the SAR-55 genome were synthesized using an Applied Biosystems model 391 DNA synthesizer.

[0078] The sequences of these 94 primers are shown below starting with SEQ. ID NO. 5 and continuing to SEQ. ID NO. 98:

HEV Primer List

Primer	ORF Region	Sequence	
D 3042 B	1	**ACATTTGAATTCACAGACAT TGTGC**	(SEQ. ID. NO. 5)
R 3043 B	1	**ACACAGATCTGAGCTACATT CGTGAG**	(SEQ. ID. NO. 6)

(continued)

Primer	ORF Region	Sequence	
D 3044 B	1	AAAGGGATCCATGGTGTTTGAGAATGZ	(SEQ. ID. NO. 7)
R 3045 B	1	ACTCACTGCAGAGCACTATCGAATC	(SEQ. ID. NO. 8)
R 261 S	1	CGGTAAACTGGTACTGCACAAC	(SEQ. ID. NO. 9)
D 260 S	1	AAGTCCCGCTCTATTACCCAAG	(SEQ. ID. NO. 10)
D 259 S	1	ACCCACGGGTGTTGGTTTTTG	(SEQ. ID. NO. 11)
R 255 S	1	TTCTTGGGGCAGGTAGAGAAG	(SEQ. ID. NO. 12)
R 254 S	2	TTATTGAATTCATGTCAACGGACGTC	(SEQ. ID. NO. 13)
D 242 S	1	AATAATTCATGCCGTCGCTCC	(SEQ. ID. NO. 14)
R 241 S	1	AAGCTCAGGAAGGTACAACTC	(SEQ. ID. NO. 15)
R 231 S	1	AAATCGATGGCTGGGATCTGATTC	(SEQ. ID. NO. 16)
R 230 S	1	GAGGCATTGTAGAGCTTTGTG	(SEQ. ID. NO. 17)
D 229 S	1	GATGTTGCACGGACAGCAAATC	(SEQ. ID. NO. 18)
D 228 S	1	ATCTCCGATGCAATCGTTAATAAC	(SEQ. ID. NO. 19)
D 227 B	1	TAATCCATTCTGTGGCGAGAG	(SEQ. ID. NO. 20)
R 218 B	2	AAGTGTGACCTTGGTCCAGTC	(SEQ. ID. NO. 21)
D 217 B	2	TTGCTCGTGCCACAATTCGCTAC	(SEQ. ID. NO. 22)

(continued)

Primer	ORF Region	Sequence	
D 211 B	1	CATTTCACTGAGTCAGTGAAGZ	(SEQ. ID. NO. 23)
D 202 B	2	TAATTATAACACCACTGCTAG	(SEQ. ID. NO. 24)
R 201 B	2	GATTGCAATACCCTTATCCTG	(SEQ. ID. NO. 25)
R 200 S	1	ATTAAACCTGTATAGGGCAGAAC	(SEQ. ID. NO. 26)
R 199 S	1	AAGTTCGATAGCCAGATTTGC	(SEQ. ID. NO. 27)
R 198 S	2	TCATGTTGGTTGTCATAATCC	(SEQ. ID. NO. 28)
R 193 B	1	GATGACGCACTTCTCAGTGTG	(SEQ. ID. NO. 29)
R 192 B	1	AGAACAACGAACGGAGAAC	(SEQ. ID. NO. 30)
D 191 B	1	AGATCCCAGCCATCGACTTTG	(SEQ. ID. NO. 31)
R 190 S	2	TAGTAGTGTAGGTGGAAATAG	(SEQ. ID. NO. 32)
D 189 B	2	GTGTGGTTATTCAGGATTATG	(SEQ. ID. NO. 33)
D 188 B	2	ACTCTGTGACCTTGGTTAATG	(SEQ. ID. NO. 34)
R 187 S	2	AACTCAAGTTCGAGGGCAAAG	(SEQ. ID. NO. 35)
D 186 S	2	CGCTTACCCTGTTTAACCTTG	(SEQ. ID. NO. 36)
D 185 B	2,3	ATCCCCTATATTCATCCAACCAAC	(SEQ. ID. NO. 37)
D 184 S	2,3	CTCCTCATGTTTCTGCCTATG	(SEQ. ID. NO. 38)
R 181 S	2	GCCAGAACGAAATGGAGATAGC	(SEQ. ID. NO. 39)

(continued)

Primer	ORF Region	Sequence	
R 180 B	1	**CTCAGACATAAAACCTAAGTC**	(SEQ. ID. NO. 40)
D 179 S	1	**TGCCCTATACAGGTTTAATCG**	(SEQ. ID. NO. 41)
D 178 B	1	ACCGGCATATACCAGGTGC	(SEQ. ID. NO. 42')
D 177 B	2	**ACATGGCTCACTCGTAAATTC**	(SEQ. ID. NO. 43)
R 174 B	1	**AACATTAGACGCGTTAACGAG**	(SEQ. ID. NO. 44)
D 173 S	1	**CTCTTTTGATGCCAGTCAGAG**	(SEQ. ID. NO. 45)
D 172 B	1	**ACCTACCCGGATGGCTCTAAGG**	(SEQ. ID. NO. 46)
R 166 B	2	**TATGGGAATTCGTGCCGTCCTGAAG (EcoRI)**	(SEQ. ID. NO. 47)
R 143 B	1	**AGTGGGAGCAGTATACCAGCG**	(SEQ. ID. NO. 48)
D 141 B	1	**CTGCTATTGAGCAGGCTGCTC**	(SEQ. ID. NO. 49)
R 142 S	1	**GGGCCATTAGTCTCTAAAACC**	(SEQ. ID. NO. 50)
D 135 B	1	GAGGTTTTCTGGAATCATC	(SEQ. ID. NO. 51)
R 134 B	1	GCATAGGTGAGACTG	(SEQ. ID. NO. 52)
R 133 B	1	AGTTACAGCCAGAAAACC	(SEQ. ID. NO. 53)
D 132 S	2,3	**CCATGGATCCTCGGCCTATTTTGCTGTTGCTCC (Bam HI)**	(SEQ. ID. NO. 54)
D 131 B	5'NC	AGGCAGACCACATATGTG	(SEQ. ID. NO. 55)
R 119 B	1	GGTGCACTCCTGACCAAGCC	(SEQ. ID. NO. 56)
D 118 B	1	ATTGGCTGCCACTTTGTTC	(SEQ. ID. NO. 57)
R 117 B	1	**ACCCTCATACGTCACCACAAC**	(SEQ. ID. NO. 58)
R 116 B	1	**GCGGTGGACCACATTAGGATTATC**	(SEQ. ID. NO. 59)
D 115 B	1	CATGATATGTCACCATCTG	(SEQ. ID. NO. 60)
D 114 B	1	GTCATCCATAACGAGCTGG	(SEQ. ID. NO. 61)

(continued)

Primer	ORF Region	Sequence	
R 112 B	2	AGCGGAATTCGAGGGGCGGC ATAAAGAACCAGG (EcoRI)	(SEQ. ID. NO. 62)
R 111 B	2	GCGCTGAATTCGGATCACAA GCTCAGAGGCTATGCC (EcoRI)	(SEQ. ID. NO. 63)
D 110 B	2	GTATAACGGATCCACATCTC CCCTTACCTC (Bam HI)	(SEQ. ID. NO. 64)
D 109 B	2	TAACCTGGATCCTTATGCCG CCCCTCTTAG (Bam HI)	(SEQ. ID. NO. 65)
D 108 B	1	AAATTGGATCCTGTGTCGGG TGGAATGAATAACATGTC (BamHI)	(SEQ. ID. NO. 66)
R 107 B	1	ATCGGCAGATCTGATAGAGC GGGGACTTGCCGGATCC	(SEQ. ID. NO. 67)
D 101 B	2	TACCCTGCCCGCGCCCATAC TTTTGATG	(SEQ. ID. NO. 68)
R 100 B	1	GGCTGAGATCTGGTTCGGGT CGCCAAGAAGGTG (Bgl II)	(SEQ. ID. NO. 69)
R 99 B	2	TACAGATCTATACAACTTAA CAGTCGG (Bgl II)	(SEQ. ID. NO. 70)
R 98 B	2	GCGGCAGATCTCACCGACAC CATTAGTAC (Bgl II)	(SEQ. ID. NO. 71)
D 97 S	1	CCGTCGGATCCCAGGGGCTG CTGTCCTG (Bam HI)	(SEQ. ID. NO. 72)
R 96 B	2	AAAGGAATTCAAGACCAGAG GTAGCCTCCTC (EcoRI)	(SEQ. ID. NO. 73)
D 95 B	2	GTTGATATGAATTCAATAAC CTCGACGG	(SEQ. ID. NO. 74)
R 94 B	3'NC	TTTGGATCCTCAGGGAGCGC GGAACGCAGAAATGAG (BamHI)	(SEQ. ID. NO. 75)
D 90 B	2	TCACTCGTGAATTCCTATAC TAATAC (EcoRI)	(SEQ. ID. NO. 76)

(continued)

Primer	ORF Region	Sequence	
R 89 B	3'NC	TTTGGATCCTCAGGGAGCGC GGAACGCAGAAATG (BamHI)	(SEQ. ID. NO. 77)
R 88 B	1	TGATAGAGCGGGACTTGCCG GATCC (BamHI)	(SEQ. ID. NO. 78)
R 87 B	1	TTGCATTAGGTTAATGAGGA TCTC	(SEQ. ID. NO. 79)
D 86 B	1	ACCTGCTTCCTTCAGCCTGC AGAAG	(SEQ. ID. NO. 80)
R 81 B	1	GCGGTGGATCCGCTCCCAGG CGTCAAAAC (BamHI)	(SEQ. ID. NO. 81)
D 80 B	1	GGGCGGATCGAATTCGAGAC CCTTCTTGG (EcoRI)	(SEQ. ID. NO. 82)
R 79 B	1	AGGATGGATCCATAAGTTAC CGATCAG (BamHI)	(SEQ. ID. NO. 83)
D 78 B	1	GGCTGGAATTCCTCTGAGGA CGCCCTCAC (EcoRI)	(SEQ. ID. NO. 84)
R 77 B	1	GCCGAAGATCTATCGGACAT AGACCTC (Bgl II)	(SEQ. ID. NO. 85)
R 76 B	2	CAGACGACGGATCCCCTTGG ATATAGCCTG (BamHI)	(SEQ. ID. NO. 86)
D 75 B	5'NC	GGCCGAATTCAGGCAGACCA CATATGTGGTCGATGCCATG (EcoRI)	(SEQ. ID. NO. 87)
D 72 B	1	GCAGGTGTGCCTGGATCCGG CAAGT (BamHI)	(SEQ. ID. NO. 88)
R 71 B	1	GTTAGAATTCCGGCCCAGCT GTGGTAGGTC (EcoRI)	(SEQ. ID. NO. 89)
D 63 B	1	CCGTCCGATTGGTCTGTATG CAGG	(SEQ. ID. NO. 90)
D 61 B	1	TACCAGTTTACTGCAGGTGT GC	(SEQ. ID. NO. 91)
D 60 B	1	CAAGCCGATGTGGACGTTGT CG	(SEQ. ID. NO. 92)

(continued)

Primer	ORF Region	Sequence	
R 59 B	2,3	GGCGCTGGGCCTGGTCACGC CAAG	(SEQ. ID. NO. 93)
D 50 B	1	GCAGAAACTAGTGTTGACCC AG	(SEQ. ID. NO. 94)
R 49 B	2	TAGGTCTACGACGTGAGGCA AC	(SEQ. ID. NO. 95)
R 48 B	1	TACAATCTTTCAGGAAGAAG G	(SEQ. ID. NO. 96)
R 47 B	1	CCCACACTCCTCCATAATAG C	(SEQ. ID. NO. 97)
D 46 B	1	GATAGTGCTTTGCAGTGAGT ACCG	(SEQ. ID. NO. 98)

[0079] The abbreviations to the left of the sequences represent the following: R and D refer to reverse and forward primers, respectively; B and S refer to sequences derived from the Burma-121 Strain of Hepatitis E and the SAR-55 Strain of Hepatitis E, respectively; 5'NC and 3'NC refer to 5 prime and 3 prime non-coding regions of the HEV genome, respectively; and 1, 2 and 3 refer to sequence derived from open reading frames 1, 2 or 3, respectively. The symbol () to the right of some sequences shown indicates insertion of an artificial restriction site into these sequences.

[0080] For cloning of PCR fragments, *Eco*RI, *Bam*HI, or *Bgl*II restriction sites preceded by 3-7 nt were added to the 5' end of primers.

[0081] *RT-PCR.* The usual 100-μl RT-PCR mixture contained template, 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, all four dNTPs (each at 0.2 mM), 50 pmol of direct primer, 50 pmol of reverse primer, 40 units of RNasin (Promega), 16 units of avian myeloblastosis virus reverse transcriptase (Promega), 4 units of AmpliTaq (Cetus), under 100 μl of light mineral oil. The mixture was incubated 1 h at 42˚C and then amplified by 35 PCR cycles; 1 min at 94˚C, 1 min at 45˚C, and 1 min at 72˚C. The PCR products were analyzed on 1% agarose gels.

[0082] *Cloning of PCR Fragments.* PCR fragments containing restriction sites at the ends were digested with *Eco*RI and *Bam*Hi or *Eco*RI and *Bgl*II restriction enzymes and cloned in E*co*RI/*Bam*HI-digested pBR322 or pGEM-3Z (Promega). Alternatively, PCR fragments were cloned into pCR1000 (Invitrogen) using the TA cloning kit (Invitrogen).

[0083] *Sequencing of PCR Fragments and Plasmids.* PCR fragments were excised from 1% agarose gels and purified by Geneclean (Bio 101, La Jolla, CA). Double-stranded PCR fragments were sequenced by using Sequenase (United States Biochemical) as described in Winship, P.R. (1984), Nucleic Acids Rev., 17:1266. Double-stranded plasmids purified through CsCl gradients were sequenced with a Sequenase kit (United States Biochemical).

[0084] *Computer Analysis of Sequences*. Nucleotide sequences of HEV strains were compared using the Genetics Computer Group (Madison, WI) software package (Devereaux, J. et al. (1984), Nucleic Acids Rev., 12:387-395, version 7.5, on a VAX 8650 computer (at the National Cancer Institute, Frederick, MD)).

EXAMPLE 2

Construction of a Recombinant Expression Vector, P63-2.

[0085] A plasmid containing the complete ORF-2 of the genome of HEV strain SAR-55, Tsarev, S.A. et al. (1992), Proc. Natl. Acad, Sci. USA, 89:559-563), was used to obtain a restriction fragment NruI-BqlII. NruI cut the HEV cDNA five nucleotides upstream of the ATG initiation codon of ORF-2. An artificial Bgl II site previously had been placed at the 3' end of HEV genome just before the poly A sequence (Tsarev, S.A. et al. (1992), Proc. Natl. Acad. Sci, USA, 89: 559-563). To insert this fragment into pBlueBac-Transfer vector (Invitrogen) a synthetic polylinker was introduced into the unique NheI site in the vector. This polylinker contained Bln I and Bgl II sites which are absent in both HEV cDNA

and pBlueBac sequences. The NruI-BglII ORF-2 fragment was inserted in Bln I-BglII pBlueBac using an adapter as shown in Fig. 1.

EXAMPLE 3

Expression of P63-2 in SF9 Insect Cells.

[0086] p63-2 and AcMNPV baculovirus DNA (Invitrogen) were cotransfected into SF9 cells (Invitrogen) by the Ca precipitation method according to the Invitrogen protocol - By following this protocol; the AcMNPV baculovirus DNA can produce a live intact baculovirus which can package p63-2 to form a recombinant baculovirus. This recombinant baculovirus was plaque-purified 4 times. The resulting recombinant baculovirus 63-2-IV-2 was used to infect SF9 cells.

[0087] *SDS-PAGE and Western blot*. Insect cells were resuspended in loading buffer (50 mM Tris-HCl, pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromphenol blue and 10% glycerol) and SDS-polyacrylamide gel electrophoresis was performed as described, Laemmli, U.K. (1970), Nature, 227:680. Gels were stained with coomassie blue or proteins were electroblotted onto BA-85 nitrocellulose filters (Schleicher & Schuell). After transfer, nitrocellulose membranes were blocked in PBS containing 10% fetal calf serum and 0.5% gelatin. As a primary antibody, hyperimmune serum of chimpanzee-1313 diluted 1:1000 was used. As a secondary antibody, phosphatase-labeled affinity-purified goat antibody to human IgG (Kirkegaard & Perry Laboratories, Inc.) diluted 1:2000 was used.
Filters were developed in Western blue stabilized substrate for alkaline phosphatase (Promega). All incubations were performed in blocking solution, and washes were with PBS with 0.05% Tween-20 (Sigma).

[0088] *Expression of HEV ORF-2.* The major protein synthesized in SF9 cells infected with recombinant baculovirus 63-2-IV-2 was a protein with an apparent molecular weight of 74 KD (Fig. 2A). This size is a little larger than that predicted for the entire ORF-2 (71 KD). The size difference could be due to glycosylation of the protein since there is at least one potential site of glycosylation (Asn-Leu-Ser) in the N-terminal part. This protein was not detected in noninfected cells or in cells infected with wild-type nonrecombinant baculovirus. In the latter case, the major protein detected was a polyhedron protein. When the same lysates were analyzed by Western blot with serum of chimp-1313 (hyperimmunized with HEV), only proteins in the recombinant cell lysate reacted and the major band was again represented by a 74 KD protein (Fig. 2B). Minor bands were also present in the Western blot. Some of them had molecular weights higher than 74 KD, which could be due to different extents of glycosylation and some had lower molecular weights, which could reflect processing and/or degradation. Serum drawn from Chimp-1313 prior to inoculation with HEV did not react with any of the proteins by Western blot.

EXAMPLE 4

Immunoelectron Microscopy of Recombinant Infected SF9 Cells.

[0089] $5x10^6$ recombinant infected SF9 cells were sonicated in CsCl (1.30 g/ml) containing 10 mM Tris-HCl, pH 7.4, 0.3% sarcosyl and centrifuged 68 h, at 40,000 rpm (SW60Ti). 50 ul of the fraction, which had the highest ELISA response and a buoyant density of 1.30 g/ml was diluted in 1 ml PBS and 5 ul of chimp-1313 hyperimmune serum was added. The hyperimmune serum was prepared by rechallenging a previously infected chimp with a second strain of hepatitis E (Mexican HEV). Samples were incubated 1 h at room temperature and then overnight at 4˚C. Immune complexes were precipitated using a SW60Ti rotor at 30,000 rpm, 42˚C, 2 h. Pellets were resuspended in distilled water, negatively stained with 3% PTA, placed on carbon grids and examined at a magnification of 40,000 in an electron microscope EM-10, Carl Zeiss, Oberkochen, Germany.

[0090] *Detection of VLPs*. Cell lysates from insect cells infected with wild-type or recombinant baculovirus 63-2-IV-2 were fractionated by CsCl density centrifugation. When fractions of the CsCl gradient from the recombinant infected insect cells were incubated with Chimp-1313 hyperimmune serum, two kinds of virus-like particles (VLP) covered with antibody were observed in the fraction with buoyant density of 1.30 g/ml: first (Figs. 3A-3A'"), antibody covered individual particles that had a size (30 nm) and morphological structure suggestive of HEV, second (Fig. 3B), antibody-coated aggregates of particles smaller than HEV (about 20 nm) but which otherwise resembled HEV. Direct EM showed the presence of a very heterogenous population of objects including some of 30 and 20 nm in diameter respectively, which looked like virus particles but, in the absence of bound antibody, could not be confirmed as HEV. A number of IEM experiments suggested that at least some of the protein(s) synthesized from the ORF-2 region of the HEV genome, had assembled into a particulate structure. It was observed that insect cells at a later stage of infection, when the proportion of smaller proteins was higher, consistently gave better results in ELISA. Therefore, unfractionated lysates of recombinant insect cells from a later stage of infection were used as antigen in ELISA in subsequent tests.

EXAMPLE 5

Detection by ELISA Based on Antigen from Insect Cells Expressing Complete ORF-2 of Anti-HEV Following Infection with Different strains of HEV.

[0091] 5x10^6 SF9 cells infected with 63-2-IV-2 virus were resuspended in 1 ml of 10 mM Tris-HCl, pH 7.5, 0.15M NaCl then were frozen and thawed 3 times. 10 ul of this suspension was dissolved in 10 ml of carbonate buffer (pH 9.6) and used to cover one flexible microtiter assay plate (Falcon). Serum samples were diluted 1:20, 1:400 and 1:8000, or 1: 100, 1:1000 and 1:10000. The same blocking and washing solutions as described for the Western blot were used in ELISA. As a secondary antibody, peroxidase-conjugated goat IgG fraction to human IgG or horse radish peroxidase-labelled goat anti-Old or anti-New World monkey immunoglobulin was used. The results were determined by measuring the optical density (O.D.) at 405 nm.

[0092] To determine if insect cell-derived antigen representing a Pakistani strain of HEV could detect anti-HEV antibody in cynomolgus monkeys infected with the Mexican strain of HEV, 3 monkeys were examined (Fig. 4). Two monkeys cyno-80A82 and cyno-9A97, were infected with feces containing the Mexico '86 HEV strain (Ticehurst, J. et al. (1992), J. Infect. Dis., 165:835-845) and the third monkey cyno-83 was infected with a second passage of the same strain. As a control, serum samples from cyno-374, infected with the Pakistani HEV strain SAR-55, were tested in the same experiment. All 3 monkeys infected with the Mexican strain seroconverted to anti-HEV. Animals from the first passage seroconverted by week 15 and from the second passage by week 5. Interestingly, the highest anti-HEV titer among the 4 animals, was found in cyno-83, inoculated with the second passage of the Mexican strain. Cynos inoculated with the first passage of the Mexican strain developed the lowest titers while those inoculated with the first passage of the Pakistani strain developed intermediate titers.

EXAMPLE 6

Specificity of Anti-HEV ELISA Based on Antigen from Insect Cells Expressing Complete ORF-2.

[0093] To estimate if the ELISA described here specifically detected anti-HEV to the exclusion of any other type of hepatitis related antibody, serum samples of chimps were analyzed, in sets of four, infected with the other known hepatitis viruses (Garci, P. et al. (1992), J. Infect. Dis., 165:1006-1011; Farci, P. et al. (1992), Science (in press); Ponzetto, A. et al. (1987) J Infect. Dis., 155: 72-77; Rizzetto; m.et al. (1981) Hepatology 1: 567-574; reference for chimps - 1413, 1373, 1442, 1551 (HAV); and for chimps - 982, 1442, 1420, 1410 (HBV); is unpublished data from Purcell et al) (Table 1). Samples of pre-inoculation and 5 week and 15 week post-inoculation sera were analyzed in HEV ELISA at serum dilutions of 1:100, 1:1000 and 1:10000. None of the sera from animals infected with HAV, HBV, HCV and HDV reacted in the ELISA for HEV antibody, but all 4 chimps inoculated with HEV developed the IgM and IgG classes of anti-HEV.

Table 1. Serological assay of anti-HEV antibody in chimpanzees infected with different hepatitis viruses (Hepatitis A, B, C, D, E)

chimp	inoculated virus	week of seroconversion for inoculated virus	preserum		weeks post-inoculation					
					5		15		20/25	
			IgG	IgM	IgG	IgM	IgG	IgM	IgG	IgM
Chimp-1413	HAV	5	-	-	-	-	-	-		
Chimp-1373	HAV	7	-	-	-	-	-	-		
Chimp-1442	HAV	5	-	-	-	-	-	-		
Chimp-1451	HAV	5	-	-	-	-	-	-		
Chimp-982	HBV	3	-	-	-	-	-	-		
Chimp-1442	HBV	7	-	-	-	-	-	-	-	-
Chimp-1420	HBV	9	-	-	-	-	-	-		

(continued)

chimp	inoculated virus	week of seroconversion for inoculated virus	preserum		weeks post-inoculation					
					5		15		20/25	
			IgG	IgM	IgG	IgM	IgG	IgM	IgG	IgM
Chimp-1410	HBV	5	-	-	-	-	-	-		
Chimp-51	HCV	10	-	-	-	-	-	-		
Chimp-502	HCV	12	-	-	-	-	-	-		
Chimp-105	HCV	28	-	-	-	-	-	-		
Chimp-793	HCV	13	-	-	-	-	-	-		
Chimp-904	HDV	8	-	-	-	-	-	-		
Chimp-814	HDV	7	-	-	-	-	-	-		
Chimp-800	HDV	10	-	-	-	-	-	-		
Chimp-29	HDV	10	-	-	-	-	-	-	-	-
Chimp-1313	HEV	5	-	-	1:10000	1:100	1:1000	-		
Chimp-1310	HEV	5	-	-	1:10000	1:100	1:10000	-		
Chimp-1374	HEV	3	-	-	1:8000	-	1:8000	-		
Chimp-1375	HEV	3	-	-	1:8000	1:400	1:400	-		
Chimp-1313	HEV1st°**	5	-	-	1:10000	1:1000	1:1000	-		
Chimp-1313	HEV2nd°**	0.5	1:100	-	1:10000	-	1:10000	-		

* Chimp-1374 was positive for IgM anti-HEV three and four weeks post-inoculation (see Fig.5)

** Chimp-1313 was inoculated with HEV twice. 1st inoculation with pooled samples of 7 Pakistani patients. 2nd inoculation 45 months later with Mexican strain of HEV.

EXAMPLE 7

Determination of the Host Range of the SAR-55 Strain of HEV in Non-Human Primates.

[0094] Different primate species were inoculated with a standard stool suspension of HEV and serial serum samples were collected to monitor for infection. Serum ALT levels were determined as an indicator of hepatitis while seroconversion was defined by detection of anti-HEV.

[0095] Both rhesus monkeys inoculated with HEV (Table 2) showed very prominent peaks of ALT activity as well as seroconversion. The first sign of increased ALT activity was observed on day 14 for both animals, and definitive seroconversion occurred on day 21. The maximum titer of anti-HEV was achieved on day 29.

Table 2. Biochemical and serological profiles of HEV infection in 8 primate species

Species	Alanine aminotransferase		Anti-HEV IgG		
	day first elevated	peak value (U/L)	day first detected	initial titer	maximum titer
Rhesus-726	14	428	21	1:20	≤1:8000
Rhesus-938	14,47*	189,165	21	1:400	≤1:8000
AGM-74	7,21*	106,141	28	1:400	≤1:8000
AGH-230	21	334	21	1:8000	≤1:8000
PTM-98	21	47	21	1:8000	≤1:8000
PTM-99	14	59	21	1:400	≤1:8000
Tam-616	no	no	no	no	no
Tam-636	no	no	no	no	no
SQM-868	6,28,47*	288,208,355	41	1:400	1:400
SQM-869	41	679	35	1:400	1:400
OWM-924	6,35,54*	53,97,65	21	1:20	1:8000
OWM-925	6,91*	77,199	21'	1:20	1:8000

*Bimodal or trimodal elevation of ALT. no = no ALT elevation or anti-HEV antibody detected.

[0096] Both African green monkeys used in this study (Table 2) developed increased ALT activity and anti-HEV. Although AGM-230 died seven weeks post-inoculation, signs of infection were observed prior to that time. AGM-74 showed a bimodal increase in ALT activity as has been reported for other species (Tsarev, S.A. et al. (1992), J. Infect. Dis. (in press)). Seroconversion for AGM-74 and AGM-230 was first observed on days 27 and 21, respectively.

[0097] Although both inoculated pigtail macaques demonstrated a slight increase in ALT activity these increases were not as prominent as in the case of the animals described previously. However, both monkeys seroconverted on day 21 and the anti-HEV titers were equivalent to those of the chimps and other old World monkeys.

[0098] Neither tamarin inoculated in this study showed any rise in ALT activity or seroconversion to anti-HEV (Table 2). The squirrel monkeys responded with significantly lower levels of anti-HEV than chimps or Old World monkeys (Table 2). The time of seroconversion was also delayed as compared with those other animals. SQM-868 seroconverted on day 41, and SQM-869 seroconverted on day 35. The anti-HEV titer was not higher than 1:400 at any time during more than three months of monitoring and clearly was waning in both animals after reaching a peak value on days 47-54. However, the increases in ALT activity were rather prominent in both animals.

[0099] The owl monkeys responded to HEV infection approximately as well as the old World monkey species (Table 2). Both OWMs seroconverted on day 21 and by day 28 the anti-HEV titer had reached a value of 1:8000. ALT activity peaked on day 35 in OWM-924, but not until day 91 in OWM-925.

EXAMPLE 8

Detection of IgM and IgG Anti-HEV in Chimps.

[0100] In both chimps, the serum ALT levels increased about 4 weeks post-inoculation (Table 2, Fig. 5). Both chimps seroconverted at the time of ALT enzyme elevation or earlier (Fig. 5A, 5C). Levels of IgM anti-HEV also were determined for the chimps. In chimp-1374, the titer of IgM anti-HEV (Fig 5B) was not as high as the IgG titer (Fig 5A) and waned over two weeks. Although both IgG and IgM antibodies were first detected for this animal on day 20, the titer of IgM anti-HEV was the highest while the titer of IgG was the lowest on that day, but then rose and stayed approximately at the same level for more than three months. In chimp-1375, only IgM anti-HEV was detected on day 20 (Fig. 5D). The titer was higher than in chimp-1374 and IgM anti-HEV was detected during the entire period of monitoring. IgG anti-HEV was first observed in this animal on day 27 (Fig. 5C) and remained at approximately the same level throughout the experiment.

EXAMPLE 9

Comparison of ELISA Based on Complete ORF-2 Protein Expressed in Insect Cells With That Based on Fragments of Structural Proteins Expressed in E. coli.

[0101] To estimate if expression of the complete ORF-2 region of the HEV genome in eukaryotic cells had any advantage over expression of fragments of structural proteins in *E. coli*, we used the former antigen in ELISA to retest cynomolgus monkey sera that had been analyzed earlier (Tsarev, S.A. et al. (1992), Proc. Natl. Acad, Sci USA, 89: 559-563; and Tsarev, S.A. et al. (1992) J. Infect. Dis. (submitted)), using the antigen fragments expressed in bacteria (Table 3).

Table 3. Comparison of ELISA based on antigen from insect cells expressing complete ORF-2 with that based on antigen from E.coli expressing fragments of structural proteins

Cyno #	antigen derived from bacterial cells (Portion of ORF-2)*	antigen derived from insect cells (Complete ORF-2)		
		anti-HEV		
	day anti-HEV first detected	first detected day	titer	max. titer
Cyno-376	28	21	1:400	1:8000
Cyno-369	54	40	1:100	1:8000
Cyno-374	19	19	1:400	1:8000
Cyno-375	26	26	1:400	1:8000
Cyno-379	21	19	1:100	1:8000
Cyno-381	28	28	1:400	1:8000

* The sera were also tested with less sensitive ORF-3 antigen [..].
Tsarev, S.A. et al. (1992), J. Infect. Dis. (in press)

For 3 of the 6 monkeys examined by ELISA, the antigen expressed in insect cells detected seroconversion earlier than the antigen expressed in *E. coli*. Using the insect cell-derived antigen, we were able to detect anti-HEV antibody in sera from all six monkeys at the highest dilution tested (1:8000). With *E. coli*-cell derived antigen (Burma Strain) no information about anti-HEV titers were obtained, since all sera were tested only at a dilution of 1:100 (Tsarev, SA et al (1992) Proc. Nat. Acad. Sci. USA; 89:559-563; Tsarev et al. (1992) J. Infect. Dis. (submitted)).

[0102] In another study, hepatitis E virus, strain SAR-55 was serially diluted in 10-fold increments and the 10^{-1} through 10^{-5} dilutions were inoculated into pairs of cynomolgus monkeys to titer the virus. The serum ALT levels were measured to determine hepatitis and serum antibody to HEV was determined by the ELISA method of the present invention (data in figures) or by Genelab's ELISA (data shown as positive (+) or negative (-) test at bottom of Figures 6 a-g). All samples were tested under code.

[0103] The ELISA method of the present invention detected seroconversion to IgG anti-HEV in all cynos inoculated and all dilutions of virus.

[0104] In contrast, Genelab's results were strikingly variable, as summarized below.

Table 4.

Dilution of Virus	Genelab's ELISA	ELISA of Present Invention
10^{-1}	did not test	positive
10^{-2}	positive for both animals, limited duration	positive
10^{-3}	negative for both animals	positive
10^{-4}	Cyno 389: positive for IgM and IgG	positive
	Cyno 383: negative	positive
10^{-5}	Cyno 386: negative	positive
	Cyno 385: positive	positive

[0105] Since Cyno 385 (10^{-5}) was positive in ELISA tests both by Genelabs and the present invention, the 10^{-4} (ten times more virus inoculated) and 10^{-3} (100 times more virus inoculated) would also have been expected to be positive.

The present invention scored them as positive in contrast to Genelab's ELISA test which missed both positives at 10^{-3} and one at 10^{-4} even though the ALT levels of Cyno 383 and 393 suggested active hepatitis. Therefore, the data support the advantages of the present ELISA method over the prior art methods of detecting antibodies to HEV.

EXAMPLE 10

Use of the Complete ORF-2 Protein as a Vaccine.

[0106] As previously described above, the recombinant ORF-2 protein is immunoreactive. Additionally, it has been shown to react with a variety of sera taken from different species of animals infected with different strains of HEV. This provides support for the use of this recombinant protein as a vaccine to protect against a variety of HEV strains. Mammals are immunized with purified or partially purified recombinant ORF-2 protein in an amount sufficient to stimulate the production of protective antibodies. The immunized animals challenged with the wild-type strain of HEV are protected.

[0107] The contents of all citations, i.e., journal articles, patents and the like, are incorporated herein by reference.

[0108] It is understood that the examples and embodiments described herein are for illustrative purposes and that various modifications and changes in light thereof to persons skilled in the art are included within the spirit and purview of this application and scope of the appended claims.

SEQUENCE LISTING

[0109]

(1) GENERAL INFORMATION:

(i) APPLICANTS: Tsarev, Sergei A., Emerson, Suzanne U., Purcell, Robert H.
(ii) TITLE OF INVENTION: Recombinant Proteins of A Pakistani Strain of Hepatitis E And Their Use In Diagnostic Methods And Vaccines
(iii) NUMBER OF SEQUENCES: 98
(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: MORGAN & FINNEGAN
(B) STREET: 345 PARK AVENUE
(C) CITY: NEW YORK
(D) STATE: NEW YORK
(E) COUNTRY: USA
(F) ZIP: 10154

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: FLOPPY DISK
(B) COMPUTER: IBM PC COMPATIBLE
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: WORDPERFECT 5.1

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 07/947,263
(B) FILING DATE: 18-SEP-1992
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Bork, Richard, W.
(B) REGISTRATION NUMBER: 36,459
(C) REFERENCE/DOCKET NUMBER: 2026-4032

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (212) 758-4800
(B) TELEFAX: (212) 751-6849

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1693 amino acid residues
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Met Glu Ala His Gln Phe Ile Lys Ala Pro Gly Ile Thr Thr Ala
1               5                  10                  15

Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser Ala Leu Ala Asn
            20                  25                  30

Ala Val Val Val Arg Pro Phe Leu Ser His Gln Gln Ile Glu Ile
        35                  40                  45

Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg Pro Glu
    50                  55                  60

Val Phe Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu Leu
65                  70                  75

Glu Leu Tyr Cys Arg Ala Arg Ser Gly Arg Cys Leu Glu Ile Gly
                80                  85                  90

Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Val His Arg
            95                  100                 105

Cys Phe Leu Arg Pro Ala Gly Arg Asp Val Gln Arg Trp Tyr Thr
        110                 115                 120

Ala Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu
    125                 130                 135

Arg Gly Leu Pro Ala Ala Asp Arg Thr Tyr Cys Phe Asp Gly Phe
140                 145                 150

Ser Gly Cys Asn Phe Pro Ala Glu Thr Gly Ile Ala Leu Tyr Ser
            155                 160                 165

Leu His Asp Met Ser Pro Ser Asp Val Ala Glu Ala Met Phe Arg
        170                 175                 180

His Gly Met Thr Arg Leu Tyr Ala Ala Leu His Leu Pro Pro Glu
    185                 190                 195

Val Leu Leu Pro Pro Gly Thr Tyr Arg Thr Ala Ser Tyr Leu Leu
200                 205                 210

Ile His Asp Gly Arg Arg Val Val Val Thr Tyr Glu Gly Asp Thr
            215                 220                 225

Ser Ala Gly Tyr Asn His Asp Val Ser Asn Leu Arg Ser Trp Ile
        230                 235                 240

Arg Thr Thr Lys Val Thr Gly Asp His Pro Leu Val Ile Glu Arg
    245                 250                 255

Val Arg Ala Ile Gly Cys His Phe Val Leu Leu Leu Thr Ala Ala
260                 265                 270

Pro Glu Pro Ser Pro Met Pro Tyr Val Pro Tyr Pro Arg Ser Thr
            275                 280                 285

Glu Val Tyr Val Arg Ser Ile Phe Gly Pro Gly Gly Thr Pro Ser
        290                 295                 300
```

Leu Phe Pro Thr Ser Cys Ser Thr Lys Ser Thr Phe His Ala Val
305 310 315

Pro Ala His Ile Trp Asp Arg Leu Met Leu Phe Gly Ala Thr Leu
320 325 330

Asp Asp Gln Ala Phe Cys Cys Ser Arg Leu Met Thr Tyr Leu Arg
335 340 345

Gly Ile Ser Tyr Lys Val Thr Val Gly Thr Leu Val Ala Asn Glu
350 355 360

Gly Trp Asn Ala Ser Glu Asp Ala Leu Thr Ala Val Ile Thr Ala
365 370 375

Ala Tyr Leu Thr Ile Cys His Gln Arg Tyr Leu Arg Thr Gln Ala
380 385 390

Ile Ser Lys Gly Met Arg Arg Leu Glu Arg Glu His Ala Gln Lys
395 400 405

Phe Ile Thr Arg Leu Tyr Ser Trp Leu Phe Glu Lys Ser Gly Arg
410 415 420

Asp Tyr Ile Pro Gly Arg Gln Leu Glu Phe Tyr Ala Gln Cys Arg
425 430 435

Arg Trp Leu Ser Ala Gly Phe His Leu Asp Pro Arg Val Leu Val
440 445 450

Phe Asp Glu Ser Ala Pro Cys His Cys Arg Thr Ala Ile Arg Lys
455 460 465

Ala Val Ser Lys Phe Cys Cys Phe Met Lys Trp Leu Gly Gln Glu
470 475 480

Cys Thr Cys Phe Leu Gln Pro Ala Glu Gly Val Val Gly Asp Gln
485 490 495

Gly His Asp Asn Glu Ala Tyr Glu Gly Ser Asp Val Asp Pro Ala
500 505 510

Glu Ser Ala Ile Ser Asp Ile Ser Gly Ser Tyr Val Val Pro Gly
515 520 525

Thr Ala Leu Gln Pro Leu Tyr Gln Ala Leu Asp Leu Pro Ala Glu
530 535 540

Ile Val Ala Arg Ala Gly Arg Leu Thr Ala Thr Val Lys Val Ser
545 550 555

Gln Val Asp Gly Arg Ile Asp Cys Glu Thr Leu Leu Gly Asn Lys
560 565 570

Thr Phe Arg Thr Ser Phe Val Asp Gly Ala Val Leu Glu Thr Asn
575 580 585

Gly Pro Glu Arg His Asn Leu Ser Phe Asp Ala Ser Gln Ser Thr
590 595 600

Met Ala Ala Gly Pro Phe Ser Leu Thr Tyr Ala Ala Ser Ala Ala
605 610 615

Gly Leu Glu Val Arg Tyr Val Ala Ala Gly Leu Asp His Arg Ala
620 625 630

Val Phe Ala Pro Gly Val Ser Pro Arg Ser Ala Pro Gly Glu Val
635 640 645

Thr Ala Phe Cys Ser Ala Leu Tyr Arg Phe Asn Arg Glu Ala Gln
650 655 660

Arg Leu Ser Leu Thr Gly Asn Phe Trp Phe His Pro Glu Gly Leu
665 670 675

Leu Gly Pro Phe Ala Pro Phe Ser Pro Gly His Val Trp Glu Ser
680 685 690

Ala Asn Pro Phe Cys Gly Glu Ser Thr Leu Tyr Thr Arg Thr Trp
695 700 705

Ser Glu Val Asp Ala Val Pro Ser Pro Ala Gln Pro Asp Leu Gly
710 715 720

Phe Thr Ser Glu Pro Ser Ile Pro Ser Arg Ala Ala Thr Pro Thr
725 730 735

Pro Ala Ala Pro Leu Pro Pro Pro Ala Pro Asp Pro Ser Pro Thr
740 745 750

Leu Ser Ala Pro Ala Arg Gly Glu Pro Ala Pro Gly Ala Thr Ala
755 760 765

Arg Ala Pro Ala Ile Thr His Gln Thr Ala Arg His Arg Arg Leu
770 775 780

Leu Phe Thr Tyr Pro Asp Gly Ser Lys Val Phe Ala Gly Ser Leu
785 790 795

Phe Glu Ser Thr Cys Thr Trp Leu Val Asn Ala Ser Asn Val Asp
800 805 810

His Arg Pro Gly Gly Gly Leu Cys His Ala Phe Tyr Gln Arg Tyr
815 820 825

Pro Ala Ser Phe Asp Ala Ala Ser Phe Val Met Arg Asp Gly Ala
830 835 840

Ala Ala Tyr Thr Leu Thr Pro Arg Pro Ile Ile His Ala Val Ala
845 850 855

Pro Asp Tyr Arg Leu Glu His Asn Pro Lys Arg Leu Glu Ala Ala
860 865 870

Tyr Arg Glu Thr Cys Ser Arg Leu Gly Thr Ala Ala Tyr Pro Leu
875 880 885

Leu Gly Thr Gly Ile Tyr Gln Val Pro Ile Gly Pro Ser Phe Asp
890 895 900

Ala Trp Glu Arg Asn His Arg Pro Gly Asp Glu Leu Tyr Leu Pro
905 910 915

Glu Leu Ala Ala Arg Trp Phe Glu Ala Asn Arg Pro Thr Cys Pro
920 925 930

Thr Leu Thr Ile Thr Glu Asp Val Ala Arg Thr Ala Asn Leu Ala
935 940 945

Ile Glu Leu Asp Ser Ala Thr Asp Val Gly Arg Ala Cys Ala Gly
950 955 960

Cys Arg Val Thr Pro Gly Val Val Gln Tyr Gln Phe Thr Ala Gly
965 970 975

Val Pro Gly Ser Gly Lys Ser Arg Ser Ile Thr Gln Ala Asp Val
980 985 990

Asp Val Val Val Val Pro Thr Arg Glu Leu Arg Asn Ala Trp Arg
995 1000 1005

Arg Arg Gly Phe Ala Ala Phe Thr Pro His Thr Ala Ala Arg Val
1010 1015 1020

Thr Gln Gly Arg Arg Val Val Ile Asp Glu Ala Pro Ser Leu Pro
1025 1030 1035

Pro His Leu Leu Leu Leu His Met Gln Arg Ala Ala Thr Val His
1040 1045 1050

Leu Leu Gly Asp Pro Asn Gln Ile Pro Ala Ile Asp Phe Glu His
1055 1060 1065

Ala Gly Leu Val Pro Ala Ile Arg Pro Asp Leu Ala Pro Thr Ser
1070 1075 1080

Trp Trp His Val Thr His Arg Cys Pro Ala Asp Val Cys Glu Leu
1085 1090 1095

Ile Arg Gly Ala Tyr Pro Met Ile Gln Thr Thr Ser Arg Val Leu
1100 1105 1110

Arg Ser Leu Phe Trp Gly Glu Pro Ala Val Gly Gln Lys Leu Val
1115 1120 1125

Phe Thr Gln Ala Ala Lys Ala Ala Asn Pro Gly Ser Val Thr Val
1130 1135 1140

His Glu Ala Gln Gly Ala Thr Tyr Thr Glu Thr Thr Ile Ile Ala
1145 1150 1155

Thr Ala Asp Ala Arg Gly Leu Ile Gln Ser Ser Arg Ala His Ala
1160 1165 1170

Ile Val Ala Leu Thr Arg His Thr Glu Lys Cys Val Ile Ile Asp
1175 1180 1185

Ala Pro Gly Leu Leu Arg Glu Val Gly Ile Ser Asp Ala Ile Val
1190 1195 1200

Asn Asn Phe Phe Leu Ala Gly Gly Glu Ile Gly His Gln Arg Pro
1205 1210 1215

Ser Val Ile Pro Arg Gly Asn Pro Asp Ala Asn Val Asp Thr Leu
1220 1225 1230

Ala Ala Phe Pro Pro Ser Cys Glu Ile Ser Ala Phe His Glu Leu
1235 1240 1245

Ala Glu Glu Leu Gly His Arg Pro Ala Pro Val Ala Ala Val Leu
1250 1255 1260

Pro Pro Cys Pro Glu Leu Glu Gln Gly Leu Leu Tyr Leu Pro Gln
1265 1270 1275

Glu Leu Thr Thr Cys Asp Ser Val Val Thr Phe Glu Leu Thr Asp
1280 1285 1290

Ile Val His Cys Arg Met Ala Ala Pro Ser Gln Arg Lys Ala Val
1295 1300 1305

Leu Ser Thr Leu Val Gly Arg Tyr Gly Arg Arg Thr Lys Leu Tyr
1310 1315 1320

Asn Ala Ser His Ser Asp Val Arg Asp Ser Leu Ala Arg Phe Ile
1325 1330 1335

Pro Ala Ile Gly Pro Val Gln Val Thr Thr Cys Glu Leu Tyr Glu
1340 1345 1350

Leu Glu Glu Ala Met Val Glu Lys Gly Gln Asp Gly Ser Ala Val
1355 1360 1365

Leu Glu Leu Asp Leu Cys Ser Arg Asp Val Ser Arg Ile Thr Phe
1370 1375 1380

Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr Ile Ala
1385 1390 1395

His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe
1400 1405 1410

Cys Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Ala Ile
1415 1420 1425

Leu Ala Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp
1430 1435 1440

Asp Thr Val Phe Ser Ala Ala Val Ala Ala Ala Lys Ala Ser Met
1445 1450 1455

Val Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn
1460 1465 1470

Phe Ser Leu Gly Leu Glu Cys Ala Ile Met Glu Glu Cys Gly Met
1475 1480 1485

Pro Gln Trp Leu Ile Arg Leu Tyr His Leu Ile Arg Ser Ala Trp
1490 1495 1500

Ile Leu Gln Ala Pro Lys Glu Ser Leu Arg Gly Phe Trp Lys Lys
1505 1510 1515

His Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr Val Trp Asn
1520 1525 1530

Met Ala Val Ile Thr His Cys Tyr Asp Phe Arg Asp Leu Gln Val
1535 1540 1545

Ala Ala Phe Lys Gly Asp Asp Ser Ile Val Leu Cys Ser Glu Tyr
1550 1555 1560

```
Arg Gln Ser Pro Gly Ala Ala Val Leu Ile Ala Gly Cys Gly Leu
1565                1570                1575

Lys Leu Lys Val Asp Phe Arg Pro Ile Gly Leu Tyr Ala Gly Val
1580                1585                1590

Val Val Ala Pro Gly Leu Gly Ala Leu Pro Asp Val Val Arg Phe
1595                1600                1605

Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro Gly Pro Glu Arg
1610                1615                1620

Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu Arg Lys Leu
1625                1630                1635

Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg Val Tyr
1640                1645                1650

Gly Val Ser Pro Gly Leu Val His Asn Leu Ile Glu Met Leu Gln
1655                1660                1665

Ala Val Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro
1670                1675                1680

Val Leu Asp Leu Thr Asn Ser Ile Leu Cys Arg Val Glu
1685                1690
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 660 amino acid residues
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Arg Pro Arg Pro Ile Leu Leu Leu Leu Leu Met Phe Leu Pro
1               5                  10                  15

Met Leu Pro Ala Pro Pro Pro Gly Gln Pro Ser Gly Arg Arg Arg
            20                  25                  30

Gly Arg Arg Ser Gly Gly Ser Gly Gly Gly Phe Trp Gly Asp Arg
        35                  40                  45

Val Asp Ser Gln Pro Phe Ala Ile Pro Tyr Ile His Pro Thr Asn
    50                  55                  60

Pro Phe Ala Pro Asp Val Thr Ala Ala Ala Gly Ala Gly Pro Arg
65                  70                  75

Val Arg Gln Pro Ala Arg Pro Leu Gly Ser Ala Trp Arg Asp Gln
            80                  85                  90

Ala Gln Arg Pro Ala Ala Ala Ser Arg Arg Arg Pro Thr Thr Ala
        95                  100                 105

Gly Ala Ala Pro Leu Thr Ala Val Ala Pro Ala His Asp Thr Pro
    110                 115                 120
```

```
Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg Gln
                125                 130                 135

Tyr Asn Leu Ser Thr Ser Pro Leu Thr Ser Ser Val Ala Thr Gly
            140                 145                 150

Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu Ser Pro Leu Leu Pro
        155                 160                 165

Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala Thr Glu Ala Ser
    170                 175                 180

Asn Tyr Ala Gln Tyr Arg Val Ala Arg Ala Thr Ile Arg Tyr Arg
185                 190                 195

Pro Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser Ile Ser
            200                 205                 210

Phe Tyr Pro Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met Asn
        215                 220                 225

Ser Ile Thr Ser Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile
    230                 235                 240

Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn
245                 250                 255

Gln Gly Trp Arg Ser Val Glu Thr Ser Gly Val Ala Glu Glu Glu
            260                 265                 270

Ala Thr Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Pro Val
        275                 280                 285

Asn Ser Tyr Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu
    290                 295                 300

Asp Phe Ala Leu Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn
305                 310                 315

Thr Asn Thr Arg Val Ser Arg Tyr Ser Ser Thr Ala Arg His Arg
            320                 325                 330

Leu Arg Arg Gly Ala Asp Gly Thr Ala Glu Leu Thr Thr Thr Ala
        335                 340                 345

Ala Thr Arg Phe Met Lys Asp Leu Tyr Phe Thr Ser Thr Asn Gly
    350                 355                 360

Val Gly Glu Ile Gly Arg Gly Ile Ala Leu Thr Leu Phe Asn Leu
365                 370                 375

Ala Asp Thr Leu Leu Gly Gly Leu Pro Thr Glu Leu Ile Ser Ser
            380                 385                 390

Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro Val Val Ser Ala Asn
        395                 400                 405

Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val Glu Asn Ala Gln
    410                 415                 420

Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp Leu Gly Glu
425                 430                 435
```

```
Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln His Glu Gln Asp
                440                 445                 450

Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val Leu
                455                 460                 465

Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr
                470                 475                 480

Asp Gln Ser Thr Tyr Gly Ser Ser Thr Gly Pro Val Tyr Val Ser
                485                 490                 495

Asp Ser Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val
                500                 505                 510

Ala Arg Ser Leu Asp Trp Thr Lys Val Thr Leu Asp Gly Arg Pro
                515                 520                 525

Leu Ser Thr Ile Gln Gln Tyr Ser Lys Thr Phe Phe Val Leu Pro
                530                 535                 540

Leu Arg Gly Lys Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala
                545                 550                 555

Gly Tyr Pro Tyr Asn Tyr Asn Thr Thr Ala Ser Asp Gln Leu Leu
                560                 565                 570

Val Glu Asn Ala Ala Gly His Arg Val Ala Ile Ser Thr Tyr Thr
                575                 580                 585

Thr Ser Leu Gly Ala Gly Pro Val Ser Ile Ser Ala Val Ala Val
                590                 595                 600

Leu Ala Pro His Ser Val Leu Ala Leu Leu Glu Asp Thr Met Asp
                605                 610                 615

Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro Glu Cys
                620                 625                 630

Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val Ala
                635                 640                 645

Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Leu
                650                 655                 660
```

2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 123 amino acid resides
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Asn Asn Met Ser Phe Ala Ala Pro Met Gly Ser Arg Pro Cys
1               5                  10                  15

Ala Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys
                20                  25                  30
```

```
Cys Pro Arg His Arg Pro Val Ser Arg Leu Ala Ala Val Val Gly
                35                  40                  45

Gly Ala Ala Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu
                50                  55                  60

Ile Leu Ser Pro Ser Gln Ser Pro Ile Phe Ile Gln Pro Thr Pro
                65                  70                  75

Ser Pro Pro Met Ser Pro Leu Arg Pro Gly Leu Asp Leu Val Phe
                80                  85                  90

Ala Asn Pro Pro Asp His Ser Ala Pro Leu Gly Val Thr Arg Pro
                95                 100                 105

Ser Ala Pro Pro Leu Pro His Val Val Asp Leu Pro Gln Leu Gly
               110                 115                 120

Pro Arg Arg
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7168 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
AGGCAGACCA CATATGTGGT CGATGCCATG GAGGCCCATC AGTTTATCAA   50
GGCTCCTGGC ATCACTACTG CTATTGAGCA GGCTGCTCTA GCAGCGGCCA  100
ACTCTGCCCT TGCGAATGCT GTGGTAGTTA GGCCTTTTCT CTCTCACCAG  150
CAGATTGAGA TCCTTATTAA CCTAATGCAA CCTCGCCAGC TTGTTTTCCG  200
CCCCGAGGTT TTCTGGAACC ATCCCATCCA GCGTGTTATC CATAATGAGC  250
TGGAGCTTTA CTGTCGCGCC CGCTCCGGCC GCTGCCTCGA AATTGGTGCC  300
CACCCCCGCT CAATAAATGA CAATCCTAAT GTGGTCCACC GTTGCTTCCT  350
CCGTCCTGCC GGGCGTGATG TTCAGCGTTG GTATACTGCC CCTACCCGCG  400
GGCCGGCTGC TAATTGCCGG CGTTCCGCGC TGCGCGGGCT CCCCGCTGCT  450
GACCGCACTT ACTGCTTCGA CGGGTTTTCT GGCTGTAACT TTCCCGCCGA  500
GACGGGCATC GCCCTCTATT CTCTCCATGA TATGTCACCA TCTGATGTCG  550
CCGAGGCTAT GTTCCGCCAT GGTATGACGC GGCTTTACGC TGCCCTCCAC  600
CTCCCGCCTG AGGTCCTGTT GCCCCCTGGC ACATACCGCA CCGCGTCGTA  650
CTTGCTGATC CATGACGGCA GGCGCGTTGT GGTGACGTAT GAGGGTGACA  700
CTAGTGCTGG TTATAACCAC GATGTTTCCA ACCTGCGCTC CTGGATTAGA  750
ACCACTAAGG TTACCGGAGA CCACCCTCTC GTCATCGAGC GGGTTAGGGC  800
CATTGGCTGC CACTTTGTCC TTTTACTCAC GGCTGCTCCG GAGCCATCAC  850
CTATGCCCTA TGTCCCTTAC CCCCGGTCTA CCGAGGTCTA TGTCCGATCG  900
```

```
ATCTTCGGCC CGGGTGGCAC CCCCTCCCTA TTTCCAACCT CATGCTCCAC  950
CAAGTCGACC TTCCATGCTG TCCCTGCCCA TATCTGGGAC CGTCTCATGT 1000
TGTTCGGGGC CACCCTAGAT GACCAAGCCT TTTGCTGCTC CCGCCTAATG 1050
ACTTACCTCC GCGGCATTAG CTACAAGGTT ACTGTGGGCA CCCTTGTTGC 1100
CAATGAAGGC TGGAACGCCT CTGAGGACGC TCTTACAGCT GTCATCACTG 1150
CCGCCTACCT TACCATCTGC CACCAGCGGT ACCTCCGCAC TCAGGCTATA 1200
TCTAAGGGGA TGCGTCGCCT GGAGCGGGAG CATGCTCAGA AGTTTATAAC 1250
ACGCCTCTAC AGTTGGCTCT TTGAGAAGTC CGGCCGTGAT TATATCCCCG 1300
GCCGTCAGTT GGAGTTCTAC GCTCAGTGTA GGCGCTGGCT CTCGGCCGGC 1350
TTTCATCTTG ACCCACGGGT GTTGGTTTTT GATGAGTCGG CCCCCTGCCA 1400
CTGTAGGACT GCGATTCGTA AGGCGGTCTC AAAGTTTTGC TGCTTTATGA 1450
AGTGGCTGGG CCAGGAGTGC ACCTGTTTTC TACAACCTGC AGAAGGCGTC 1500
GTTGGCGACC AGGGCCATGA CAACGAGGCC TATGAGGGGT CTGATGTTGA 1550
CCCTGCTGAA TCCGCTATTA GTGACATATC TGGGTCCTAC GTAGTCCCTG 1600
GCACTGCCCT CCAACCGCTT TACCAAGCCC TTGACCTCCC CGCTGAGATT 1650
GTGGCTCGTG CAGGCCGGCT GACCGCCACA GTAAAGGTCT CCCAGGTCGA 1700
CGGGCGGATC GATTGTGAGA CCCTTCTCGG TAATAAAACC TTCCGCACGT 1750
CGTTTGTTGA CGGGGCGGTT TTAGAGACTA ATGGCCCAGA GCGCCACAAT 1800
CTCTCTTTTG ATGCCAGTCA GAGCACTATG GCCGCCGGCC CTTTCAGTCT 1850
CACCTATGCC GCCTCTGCTG CTGGGCTGGA GGTGCGCTAT GTCGCCGCCG 1900
GGCTTGACCA CCGGGCGGTT TTTGCCCCCG GCGTTTCACC CCGGTCAGCC 1950
CCTGGCGAGG TCACCGCCTT CTGTTCTGCC CTATACAGGT TTAATCGCGA 2000
GGCCCAGCGC CTTTCGCTGA CCGGTAATTT TTGGTTCCAT CCTGAGGGGC 2050
TCCTTGGCCC CTTTGCCCCG TTTTCCCCCG GGCATGTTTG GGAGTCGGCT 2100
AATCCATTCT GTGGCGAGAG CACACTTTAC ACCCGCACTT GGTCGGAGGT 2150
TGATGCTGTT CCTAGTCCAG CCCAGCCCGA CTTAGGTTTT ACATCTGAGC 2200
CTTCTATACC TAGTAGGGCC GCCACACCTA CCCCGGCGGC CCCTCTACCC 2250
CCCCCTGCAC CGGATCCTTC CCCTACTCTC TCTGCTCCGG CGCGTGGTGA 2300
GCCGGCTCCT GGCGCTACCG CCAGGGCCCC AGCCATAACC CACCAGACGG 2350
CCCGGCATCG CCGCCTGCTC TTTACCTACC CGGATGGCTC TAAGGTGTTC 2400
GCCGGCTCGC TGTTTGAGTC GACATGTACC TGGCTCGTTA ACGCGTCTAA 2450
TGTTGACCAC CGCCCTGGCG GTGGGCTCTG TCATGCATTT TACCAGAGGT 2500
ACCCCGCCTC CTTTGATGCT GCCTCTTTTG TGATGCGCGA CGGCGCGGCC 2550
GCCTACACAT TAACCCCCCG GCCAATAATT CATGCCGTCG CTCCTGATTA 2600
TAGGTTGGAA CATAACCCAA AGAGGCTTGA GGCTGCCTAC CGGGAGACTT 2650
GCTCCCGCCT CGGTACCGCT GCATACCCAC TCCTCGGGAC CGGCATATAC 2700
```

```
CAGGTGCCGA TCGGTCCCAG TTTTGACGCC TGGGAGCGGA ATCACCGCCC 2750
CGGGGACGAG TTGTACCTTC CTGAGCTTGC TGCCAGATGG TTCGAGGCCA 2800
ATAGGCCGAC CTGCCCAACT CTCACTATAA CTGAGGATGT TGCGCGGACA 2850
GCAAATCTGG CTATCGAACT TGACTCAGCC ACAGACGTCG GCCGGGCCTG 2900
TGCCGGCTGT CGAGTCACCC CCGGCGTTGT GCAGTACCAG TTTACCGCAG 2950
GTGTGCCTGG ATCCGGCAAG TCCCGCTCTA TTACCCAAGC CGACGTGGAC 3000
GTTGTCGTGG TCCCGACCCG GGAGTTGCGT AATGCCTGGC GCCGCCGCGG 3050
CTTCGCTGCT TTCACCCCGC ACACTGCGGC TAGAGTCACC CAGGGGCGCC 3100
GGGTTGTCAT TGATGAGGCC CCGTCCCTTC CCCCTCATTT GCTGCTGCTC 3150
CACATGCAGC GGGCCGCCAC CGTCCACCTT CTTGGCGACC CGAATCAGAT 3200
CCCAGCCATC GATTTTGAGC ACGCCGGGCT CGTTCCCGCC ATCAGGCCCG 3250
ATTTGGCCCC CACCTCCTGG TGGCATGTTA CCCATCGCTG CCCTGCGGAT 3300
GTATGTGAGC TAATCCGCGG CGCATACCCT ATGATTCAGA CCACTAGTCG 3350
GGTCCTCCGG TCGTTGTTCT GGGGTGAGCC CGCCGTTGGG CAGAAGCTAG 3400
TGTTCACCCA GGCGGCTAAG GCCGCCAACC CCGGTTCAGT GACGGTCCAT 3450
GAGGCACAGG GCGCTACCTA CACAGAGACT ACCATCATTG CCACGGCAGA 3500
TGCTCGAGGC CTCATTCAGT CGTCCCGAGC TCATGCCATT GTTGCCTTGA 3550
CGCGCCACAC TGAGAAGTGC GTCATCATTG ACGCACCAGG CCTGCTTCGC 3600
GAGGTGGGCA TCTCCGATGC AATCGTTAAT AACTTTTTCC TTGCTGGTGG 3650
CGAAATTGGC CACCAGCGCC CATCTGTTAT CCCTCGCGGC AATCCTGACG 3700
CCAATGTTGA CACCTTGGCT GCCTTCCCGC CGTCTTGCCA GATTAGCGCC 3750
TTCCATCAGT TGGCTGAGGA GCTTGGCCAC AGACCTGCCC CTGTCGCGGC 3800
TGTTCTACCG CCCTGCCCTG AGCTTGAACA GGGCCTTCTC TACCTGCCCC 3850
AAGAACTCAC CACCTGTGAT AGTGTCGTAA CATTTGAATT AACAGATATT 3900
GTGCATTGTC GTATGGCCGC CCCGAGCCAG CGCAAGGCCG TGCTGTCCAC 3950
GCTCGTGGGC CGTTATGGCC GCCGCACAAA GCTCTACAAT GCCTCCCACT 4000
CTGATGTTCG CGACTCTCTC GCCCGTTTTA TCCCGGCCAT TGGCCCCGTA 4050
CAGGTTACAA CCTGTGAATT GTACGAGCTA GTGGAGGCCA TGGTCGAGAA 4100
GGGCCAGGAC GGCTCCGCCG TCCTTGAGCT CGACCTTTGT AGCCGCGACG 4150
TGTCCAGGAT CACCTTCTTC CAGAAAGATT GTAATAAATT CACCACGGGG 4200
GAGACCATCG CCCATGGTAA AGTGGGCCAG GGCATTTCGG CCTGGAGTAA 4250
GACCTTCTGT GCCCTTTTCG GCCCCTGGTT CCGTGCTATT GAGAAGGCTA 4300
TCCTGGCCCT GCTCCCTCAG GGTGTGTTTT ATGGGGATGC CTTTGATGAC 4350
ACCGTCTTCT CGGCGGCTGT GGCCGCAGCA AAGGCATCCA TGGTGTTCGA 4400
GAATGACTTT TCTGAGTTTG ATTCCACCCA GAATAATTTT TCCTTGGGCC 4450
TAGAGTGTGC TATTATGGAG GAGTGTGGGA TGCCGCAGTG GCTCATCCGC 4500
```

```
TTGTACCACC TTATAAGGTC TGCGTGGATT CTGCAGGCCC CGAAGGAGTC  4550
CCTGCGAGGG TTTTGGAAGA AACACTCCGG TGAGCCCGGC ACCCTTCTGT  4600
GGAATACTGT CTGGAACATG GCCGTTATCA CCCACTGTTA TGATTTCCGC  4650
GATCTGCAGG TGGCTGCCTT TAAAGGTGAT GATTCGATAG TGCTTTGCAG  4700
TGAGTACCGT CAGAGCCCAG GGGCTGCTGT CCTGATTGCT GGCTGTGGCC  4750
TAAAGTTGAA GGTGGATTTC CGTCCGATTG GTCTGTATGC AGGTGTTGTG  4800
GTGGCCCCCG GCCTTGGCGC GCTTCCTGAT GTCGTGCGCT TCGCCGGTCG  4850
GCTTACTGAG AAGAATTGGG GCCCTGGCCC CGAGCGGGCG GAGCAGCTCC  4900
GCCTCGCTGT GAGTGATTTT CTCCGCAAGC TCACGAATGT AGCTCAGATG  4950
TGTGTGGATG TTGTCTCTCG TGTTTATGGG GTTTCCCCTG GGCTCGTTCA  5000
TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG GCTCATTTCA  5050
CTGAGTCAGT GAAGCCAGTG CTTGACCTGA CAAATTCAAT TCTGTGTCGG  5100
GTGGAATGAA TAACATGTCT TTTGCTGCGC CCATGGGTTC GCGACCATGC  5150
GCCCTCGGCC TATTTTGCTG TTGCTCCTCA TGTTTCTGCC TATGCTGCCC  5200
GCGCCACCGC CCGGTCAGCC GTCTGGCCGC CGTCGTGGGC GGCGCAGCGG  5250
CGGTTCCGGC GGTGGTTTCT GGGGTGACCG GGTTGATTCT CAGCCCTTCG  5300
CAATCCCCTA TATTCATCCA ACCAACCCCT TCGCCCCCGA TGTCACCGCT  5350
GCGGCCGGGG CTGGACCTCG TGTTCGCCAA CCCGCCCGAC CACTCGGCTC  5400
CGCTTGGCGT GACCAGGCCC AGCGCCCCGC CGCTGCCTCA CGTCGTAGAC  5450
CTACCACAGC TGGGGCCGCG CCGCTAACCG CGGTCGCTCC GGCCCATGAC  5500
ACCCCGCCAG TGCCTGATGT TGACTCCCGC GGCGCCATCC TGCGCCGGCA  5550
GTATAACCTA TCAACATCTC CCCTCACCTC TTCCGTGGCC ACCGGCACAA  5600
ATTTGGTTCT TTACGCCGCT CCTCTTAGCC CGCTTCTACC CCTCCAGGAC  5650
GGCACCAATA CTCATATAAT GGCTACAGAA GCTTCTAATT ATGCCCAGTA  5700
CCGGGTTGCT CGTGCCACAA TTCGCTACCG CCCGCTGGTC CCCAACGCTG  5750
TTGGTGGCTA CGCTATCTCC ATTTCGTTCT GGCCACAGAC CACCACCACC  5800
CCGACGTCCG TTGACATGAA TTCAATAACC TCGACGGATG TCCGTATTTT  5850
AGTCCAGCCC GGCATAGCCT CCGAGCTTGT TATTCCAAGT GAGCGCCTAC  5900
ACTATCGCAA CCAAGGTTGG CGCTCTGTTG AGACCTCCGG GGTGGCGGAG  5950
GAGGAGGCCA CCTCTGGTCT TGTCATGCTC TGCATACATG GCTCACCTGT  6000
AAATTCTTAT ACTAATACAC CCTATACCGG TGCCCTCGGG CTGTTGGACT  6050
TTGCCCTCGA ACTTGAGTTC CGCAACCTCA CCCCCGGTAA TACCAATACG  6100
CGGGTCTCGC GTTACTCCAG CACTGCCCGT CACCGCCTTC GTCGCGGTGC  6150
AGATGGGACT GCCGAGCTCA CCACCACGGC TGCTACTCGC TTCATGAAGG  6200
ACCTCTATTT TACTAGTACT AATGGTGTTG GTGAGATCGG CCGCGGGATA  6250
GCGCTTACCC TGTTTAACCT TGCTGACACC CTGCTTGGCG GTCTACCGAC  6300
```

```
AGAATTGATT TCGTCGGCTG GTGGCCAGCT GTTCTACTCT CGCCCCGTCG  6350
TCTCAGCCAA TGGCGAGCCG ACTGTTAAGC TGTATACATC TGTGGAGAAT  6400
GCTCAGCAGG ATAAGGGTAT TGCAATCCCG CATGACATCG ACCTCGGGGA  6450
ATCCCGTGTA GTTATTCAGG ATTATGACAA CCAACATGAG CAGGACCGAC  6500
CGACACCTTC CCCAGCCCCA TCGCGTCCTT TTTCTGTCCT CCGAGCTAAC  6550
GATGTGCTTT GGCTTTCTCT CACCGCTGCC GAGTATGACC AGTCCACTTA  6600
CGGCTCTTCG ACCGGCCCAG TCTATGTCTC TGACTCTGTG ACCTTGGTTA  6650
ATGTTGCGAC CGGCGCGCAG GCCGTTGCCC GGTCACTCGA CTGGACCAAG  6700
GTCACACTTG ATGGTCGCCC CCTTTCCACC ATCCAGCAGT ATTCAAAGAC  6750
CTTCTTTGTC CTGCCGCTCC GCGGTAAGCT CTCCTTTTGG GAGGCAGGAA  6800
CTACTAAAGC CGGGTACCCT TATAATTATA ACACCACTGC TAGTGACCAA  6850
CTGCTCGTTG AGAATGCCGC TGGGCATCGG GTTGCTATTT CCACCTACAC  6900
TACTAGCCTG GGTGCTGGCC CCGTCTCTAT TTCCGCGGTT GCTGTTTTAG  6950
CCCCCCACTC TGTGCTAGCA TTGCTTGAGG ATACCATGGA CTACCCTGCC  7000
CGCGCCCATA CTTTCGATGA CTTCTGCCCG GAGTGCCGCC CCCTTGGCCT  7050
CCAGGGTTGT GCTTTTCAGT CTACTGTCGC TGAGCTTCAG CGCCTTAAGA  7100
TGAAGGTGGG TAAAACTCGG GAGTTATAGT TTATTTGCTT GTGCCCCCCT  7150
TCTTTCTGTT GCTTATTT                                     7168
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
ACATTTGAAT TCACAGACAT TGTGC 25

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
ACACAGATCT GAGCTACATT CGTGAG 26

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
AAAGGGATCC ATGGTGTTTG AGAATG 26

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
ACTCACTGCA GAGCACTATC GAATC 25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
CGGTAAACTG GTACTGCACA AC 22

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
AAGTCCCGCT CTATTACCCA AG 22

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
ACCCACGGGT GTTGGTTTTT G 21

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
TTCTTGGGGC AGGTAGAGAA G 21

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
TTATTGAATT CATGTCAACG GACGTC 26

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
AATAATTCAT GCCGTCGCTC C 21

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
AAGCTCAGGA AGGTACAACT C 21

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
AAATCGATGG CTGGGATCTG ATTC 24

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
GAGGCATTGT AGAGCTTTGT G 21

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
GATGTTGCAC GGACAGCAAA TC 22

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
ATCTCCGATG CAATCGTTAA TAAC 24

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
TAATCCATTC TGTGGCGAGA G 21

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
AAGTGTGACC TTGGTCCAGT C 21

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:
TTGCTCGTGC CACAATTCGC TAC 23

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
CATTTCACTG AGTCAGTGAA G 21

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
TAATTATAAC ACCACTGCTA G 21

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
GATTGCAATA CCCTTATCCT G 21

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATTAAACCTG TATAGGGCAG AAC 23

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
AAGTTCGATA GCCAGATTTG C 21

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
TCATGTTGGT TGTCATAATC C 21

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
GATGACGCAC TTCTCAGTGT G 21

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:
AGAACAACGA ACGGAGAAC 19

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:
AGATCCCAGC CATCGACTTT G 21

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:
TAGTAGTGTA GGTGGAAATA G 21

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:
GTGTGGTTAT TCAGGATTAT G 21

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:
ACTCTGTGAC CTTGGTTAAT G 21

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:
AACTCAAGTT CGAGGGCAAA G 21

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:
CGCTTACCCT GTTTAACCTT G 21

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:
ATCCCCTATA TTCATCCAAC CAAC 24

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
CTCCTCATGT TTCTGCCTAT G 21

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
GCCAGAACGA AATGGAGATA GC 22

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
CTCAGACATA AAACCTAAGT C 21

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
TGCCCTATAC AGGTTTAATC G 21

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
ACCGGCATAT ACCAGGTGC 19

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
ACATGGCTCA CTCGTAAATT C 21

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
AACATTAGAC GCGTTAACGA G 21

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
CTCTTTTGAT GCCAGTCAGA G 21

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:
ACCTACCCGG ATGGCTCTAA GG 22

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
TATGGGAATT CGTGCCGTCC TGAAG 25

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
AGTGGGAGCA GTATACCAGC G 21

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:
CTGCTATTGA GCAGGCTGCT C 21

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:
GGGCCATTAG TCTCTAAAAC C 21

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:
GAGGTTTTCT GGAATCATC 19

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:
GCATAGGTGA GACTG 15

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:
AGTTACAGCC AGAAAACC 18

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:
CCATGGATCC TCGGCCTATT TTGCTGTTGC TCC 33

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
AGGCAGACCA CATATGTG 18

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
GGTGCACTCC TGACCAAGCC 20

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:
ATTGGCTGCC ACTTTGTTC 19

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:
ACCCTCATAC GTCACCACAA C 21

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
GCGGTGGACC ACATTAGGAT TATC 24

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CATGATATGT CACCATCTG 19

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:
GTCATCCATA ACGAGCTGG 19

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:
AGCGGAATTC GAGGGGCGGC ATAAAGAACC AGG 33

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:
GCGCTGAATT CGGATCACAA GCTCAGAGGC TATGCC 36

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:
GTATAACGGA TCCACATCTC CCCTTACCTC 30

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:
TAACCTGGAT CCTTATGCCG CCCCTCTTAG 30

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:
AAATTGGATC CTGTGTCGGG TGGAATGAAT AACATGTC 38

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:
ATCGGCAGAT CTGATAGAGC GGGGACTTGC CGGATCC 37

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:
TACCCTGCCC GCGCCCATAC TTTTGATG 28

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:
GGCTGAGATC TGGTTCGGGT CGCCAAGAAG GTG 33

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:
TACAGATCTA TACAACTTAA CAGTCGG 27

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:
GCGGCAGATC TCACCGACAC CATTAGTAC 29

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:
CCGTCGGATC CCAGGGGCTG CTGTCCTG 28

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:
AAAGGAATTC AAGACCAGAG GTAGCCTCCT C 31

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:
GTTGATATGA ATTCAATAAC CTCGACGG 28

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:
TTTGGATCCT CAGGGAGCGC GGAACGCAGA AATGAG 36

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:
TCACTCGTGA ATTCCTATAC TAATAC 26

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:
TTTGGATCCT CAGGGAGCGC GGAACGCAGA AATG 34

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:
TGATAGAGCG GGACTTGCCG GATCC 25

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:
TTGCATTAGG TTAATGAGGA TCTC 24

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:
ACCTGCTTCC TTCAGCCTGC AGAAG 25

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:
GCGGTGGATC CGCTCCCAGG CGTCAAAAC 29

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:
GGGCGGATCG AATTCGAGAC CCTTCTTGG 29

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:
AGGATGGATC CATAAGTTAC CGATCAG 27

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:
GGCTGGAATT CCTCTGAGGA CGCCCTCAC 29

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:
GCCGAAGATC TATCGGACAT AGACCTC 27

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:
CAGACGACGG ATCCCCTTGG ATATAGCCTG 30

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:
GGCCGAATTC AGGCAGACCA CATATGTGGT CGATGCCATG 40

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:
GCAGGTGTGC CTGGATCCGG CAAGT 25

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:
GTTAGAATTC CGGCCCAGCT GTGGTAGGTC 30

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:
CCGTCCGATT GGTCTGTATG CAGG 24

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:
TACCAGTTTA CTGCAGGTGT GC 22

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:
CAAGCCGATG TGGACGTTGT CG 22

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:
GGCGCTGGGC CTGGTCACGC CAAG 24

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GCAGAAACTA GTGTTGACCC AG 22

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:
TAGGTCTACG ACGTGAGGCA AC 22

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 basle pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:
TACAATCTTT CAGGAAGAAG G 21

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:
CCCACACTCC TCCATAATAG C 21

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:
GATAGTGCTT TGCAGTGAGT ACCG 24

Claims

1. A method for producing an immunogenic hepatitis protein comprising:

culturing an insect host cell transformed or transfected with a baculovirus expression vector, said expression vector comprising DNA, said DNA having a sequence, said sequence consisting of the nucleotides which encode the amino acids 1 to 660 of a hepatitis E virus open reading frame 2 protein, under conditions appropriate to

cause production of the immunogenic hepatitis protein.

2. The method of Claim 1, wherein said DNA sequence consists of:

a) nucleotides 5147 to 7126 of SEQ ID NO: 4,
b) a DNA sequence which encodes the amino acid sequence of SEQ ID NO: 2, or
c) a DNA sequence which encodes an amino acid sequence that is in excess of 90% homologous with the amino acid sequence of SEQ ID NO:2.

3. A baculovirus expression vector comprising DNA, said DNA having a sequence, said sequence consisting of the nucleotides which encode the amino acids 1 to 660 of a hepatitis E virus open reading frame 2 protein.

4. The expression vector of Claim 3, wherein said DNA sequence consists of:

a) nucleotides 5147 to 7126 of SEQ ID NO: 4,
b) a DNA sequence which encodes the amino acid sequence of SEQ ID NO: 2 , or
c) a DNA sequence which encodes an amino acid sequence that is in excess of 90% homologous with the amino acid sequence of SEQ ID NO: 2.

5. An insect host cell comprising the expression vector of Claim 3 or 4.

Patentansprüche

1. Verfahren zur Herstellung eines immunogenen Hepatitis-Proteins, umfassend:

Züchten einer Insekten-Wirtszelle, die mit einem Baculovirus-Expressionsvektor transformiert oder transfiziert ist, wobei der Expressionsvektor eine DNA umfasst, die eine Sequenz hat, bestehend aus den Nucleotiden, welche die Aminosäuren 1 bis 660 eines offenen Leserahmen 2-Proteins vom Hepatitis E-Virus codieren, unter Bedingungen, die zur Herstellung des immunogenen Hepatitis-Proteins geeignet sind.

2. Verfahren nach Anspruch 1, wobei die DNA-Sequenz besteht aus:

(a) Nucleotiden 5147 bis 7126 von SEQ ID NO: 4,
(b) einer DNA-Sequenz, welche die Aminosäuresequenz von SEQ ID NO: 2 codiert, oder
(c) einer DNA-Sequenz, die eine Aminosäuresequenz codiert, die zu mehr als 90 % zur Aminosäuresequenz von SEQ ID NO: 2 homolog ist.

3. Baculovirus-Expressionsvektor, umfassend DNA, wobei die DNA eine Sequenz hat, die aus den Nucleotiden besteht, welche die Aminosäuren 1 bis 660 eines offenen Leserahmen 2-Proteins vom Hepatitis E-Virus codieren.

4. Expressionsvektor nach Anspruch 3, wobei die DNA-Sequenz besteht aus:

(a) Nucleotiden 5147 bis 7126 von SEQ ID NO: 4,
(b) einer DNA-Sequenz, welche die Aminosäuresequenz von SEQ ID NO: 2 codiert, oder
(c) einer DNA-Sequenz, die eine Aminosäuresequenz codiert, die zu mehr als 90 % zur Aminosäuresequenz von SEQ ID NO: 2 homolog ist.

5. Insekten-Wirtszelle, umfassend den Expressionsvektor nach Anspruch 3 oder 4.

Revendications

1. Méthode pour produire une protéine d'hépatite immunogénique, comprenant:

la culture d'une cellule hôte d'insecte transformée ou transfectée avec un vecteur d'expression de baculovirus, ledit vecteur d'expression comprenant de l'ADN, ledit ADN ayant une séquence, ladite séquence consistant en les nucléotides qui codent les acides aminés 1 à 660 d'une protéine de cadre ouvert de lecture 2 de virus de

l'hépatite E, dans des conditions appropriées pour provoquer la production de la protéine d'hépatite immunogénique.

2. Méthode de la revendication 1, dans laquelle ladite séquence d'ADN consiste en:

a) les nucléotides 5147 à 7126 de SEQ ID NO: 4,
b) une séquence d'ADN qui code la séquence d'acides aminés de SEQ ID NO: 2, ou
c) une séquence d'ADN qui code une séquence d'acides aminés qui a plus de 90% d'homologie avec la séquence d'acides aminés de SEQ ID NO : 2.

3. Vecteur d'expression de baculovirus comprenant de l'ADN, ledit ADN ayant une séquence, ladite séquence consistant en les nucléotides qui codent les acides aminés 1 à 660 d'une protéine de cadre ouvert de lecture 2 de virus de l'hépatite E.

4. Vecteur d'expression de la revendication 3, dans lequel ladite séquence d'ADN consiste en:

a) les nucléotides 5147 à 7126 de SEQ ID NO: 4,
b) une séquence d'ADN qui code la séquence d'acides aminés de SEQ ID NO: 2, ou
c) une séquence d'ADN qui code une séquence d'acides aminés qui a plus de 90% d'homologie avec la séquence d'acides aminés de SEQ ID NO: 2.

5. Cellule hôte d'insecte comprenant le vecteur d'expression de la revendication 3 ou 4.

FIG. 1

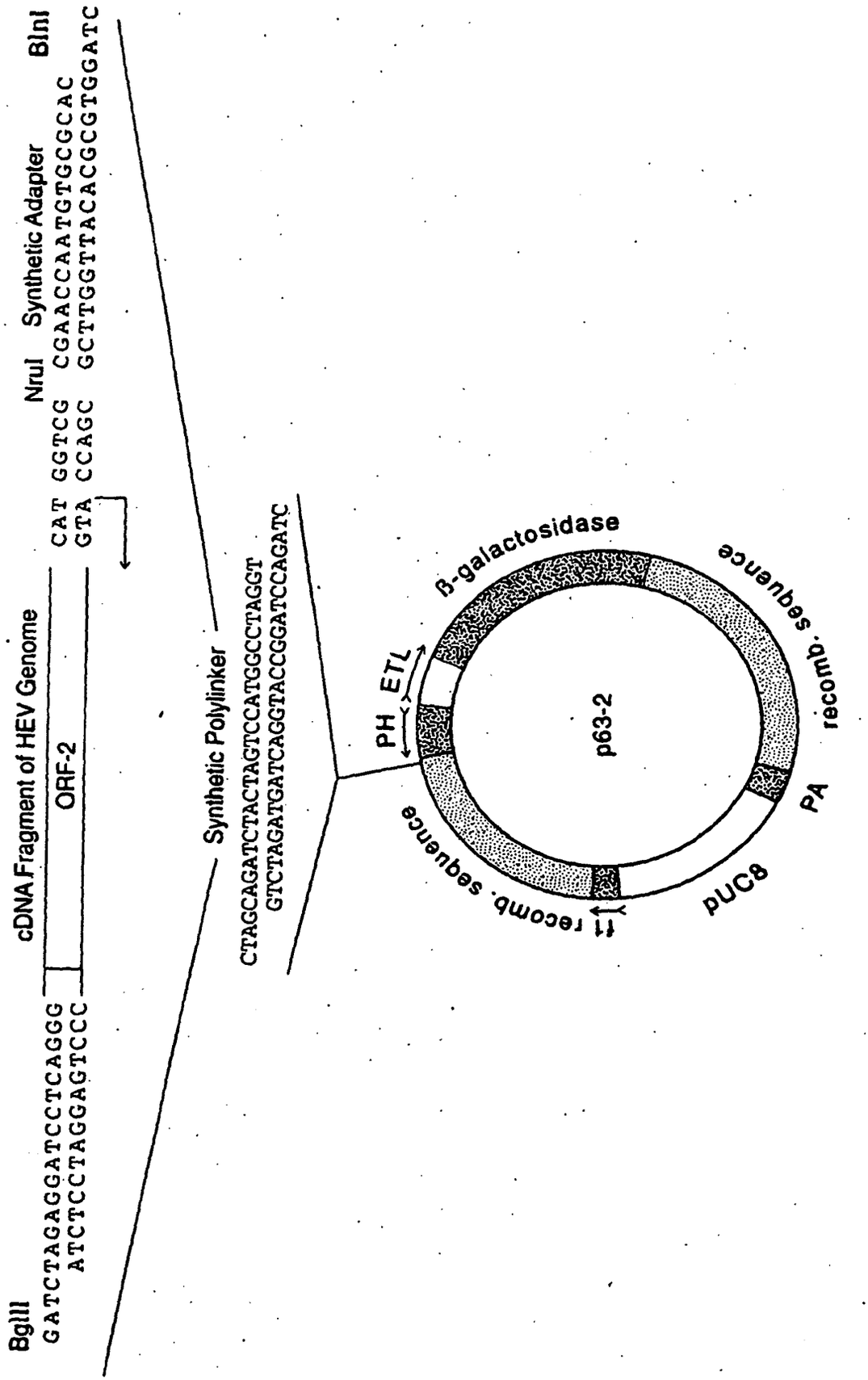

BglII
GATCTAGAGGATCCTCAGGG
ATCTCCTAGGAGTCCC
cDNA Fragment of HEV Genome
ORF-2
Synthetic Polylinker
CTAGCAGATCTACTAGTCCATGGCCTAGGT
GTCTAGATGATCAGGTACCGGATCCAGATC
NruI Synthetic Adapter BlnI
CAT GGTCG CGAACCAATGTGCGCAC
GTA CCAGC GCTTGGTTACACGCGTGGATC
PH ETL
β-galactosidase
p63-2
recomb. sequence
pUC8
PA
recomb. sequence

FIG. 2B

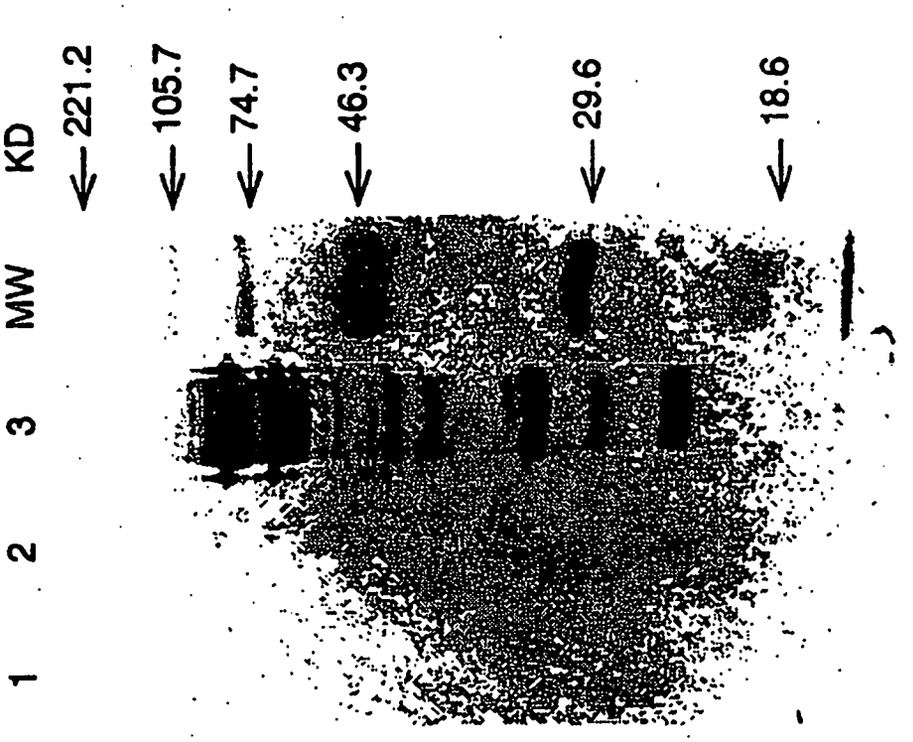

1
2
3
MW
KD
221.2
105.7
74.7
46.3
29.6
18.6

FIG. 2A

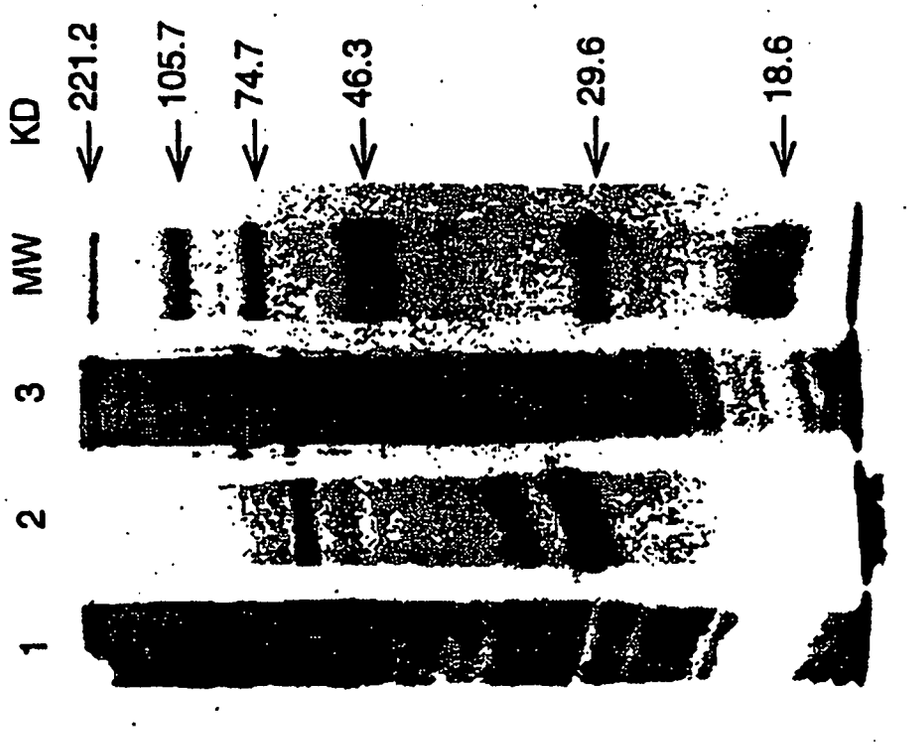

1
2
3
MW
KD
221.2
105.7
74.7
46.3
29.6
18.6

FIG. 3A

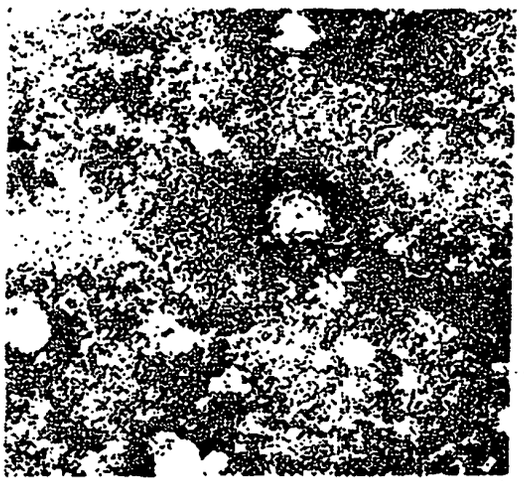


FIG. 3A'

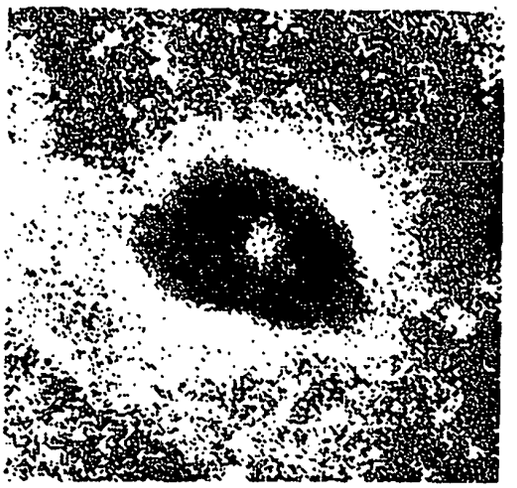


FIG. 3A''



FIG. 3A'''

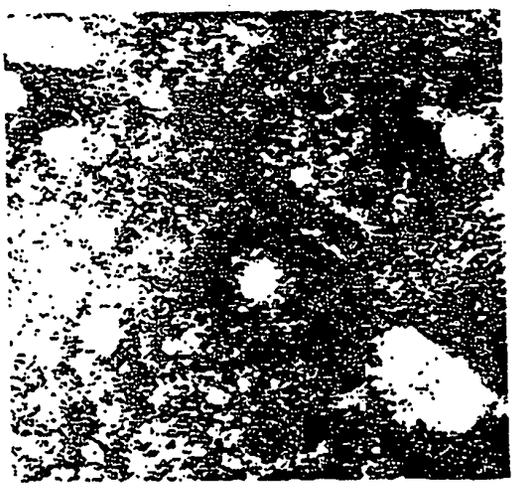


FIG. 3B

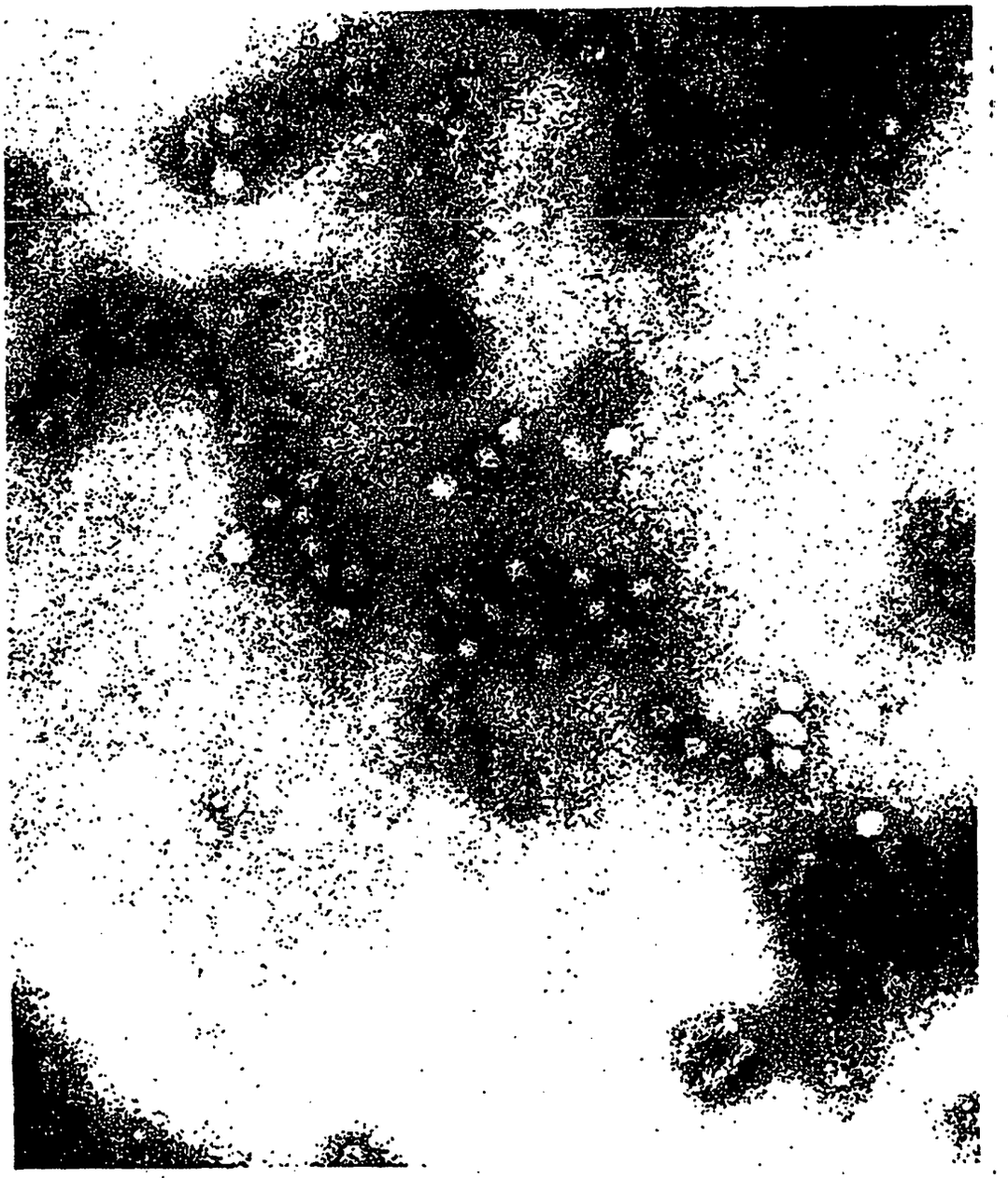


FIG. 4

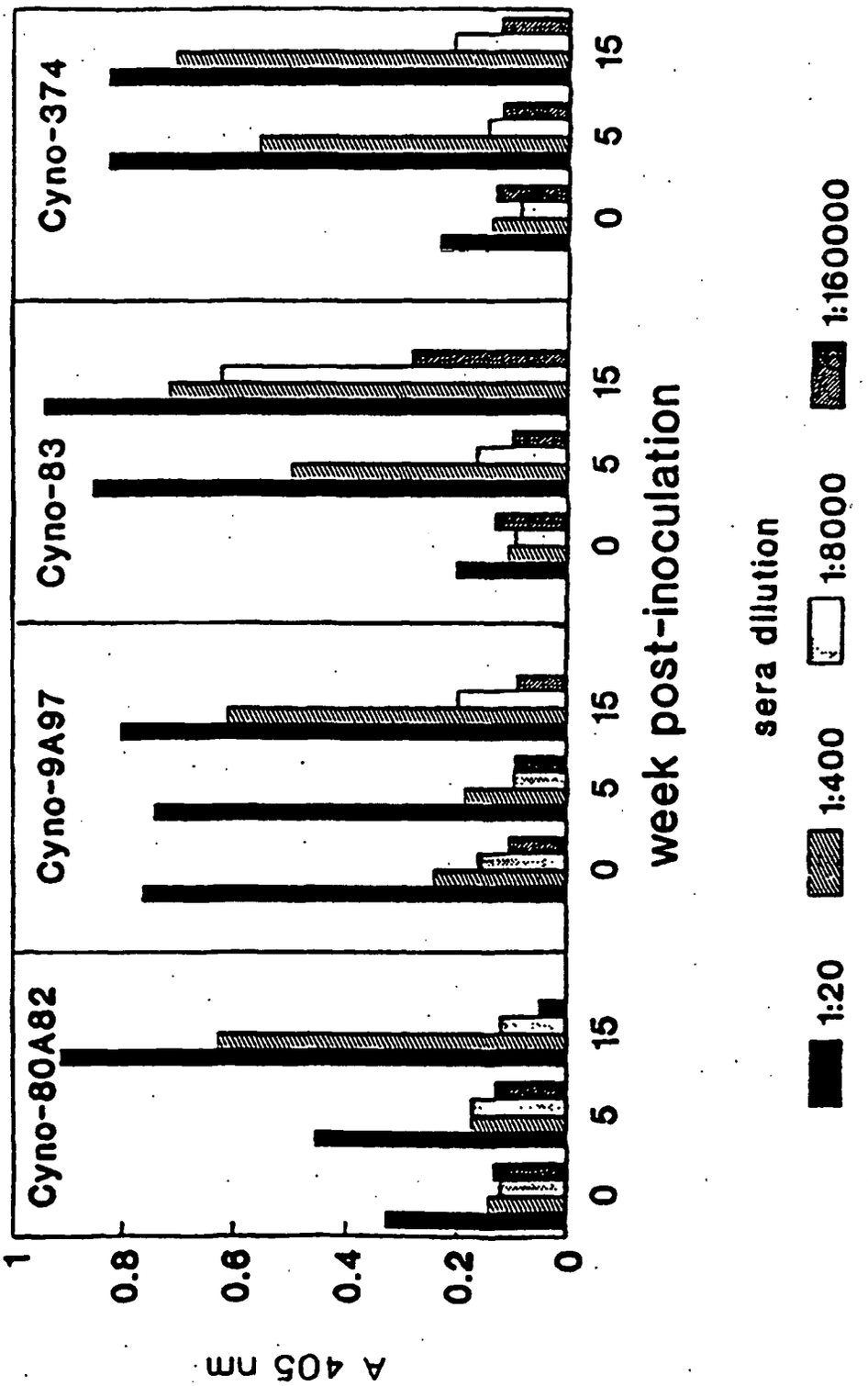

Cyno-80A82
Cyno-9A97
Cyno-83
Cyno-374
A 405 nm
1
0.8
0.6
0.4
0.2
0
0
5
15
week post-inoculation
sera dilution
1:20
1:400
1:8000
1:160000

FIG. 5A

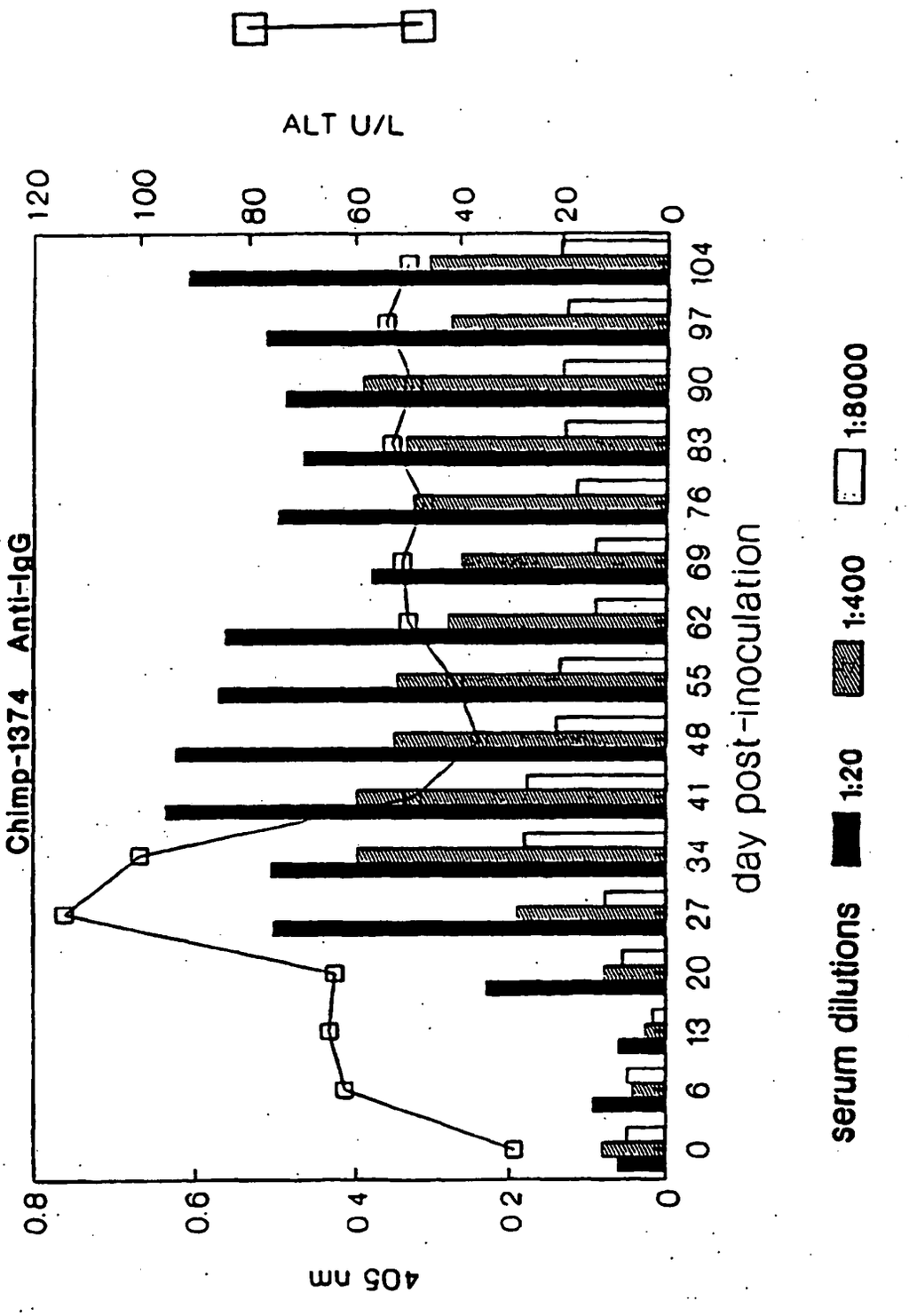
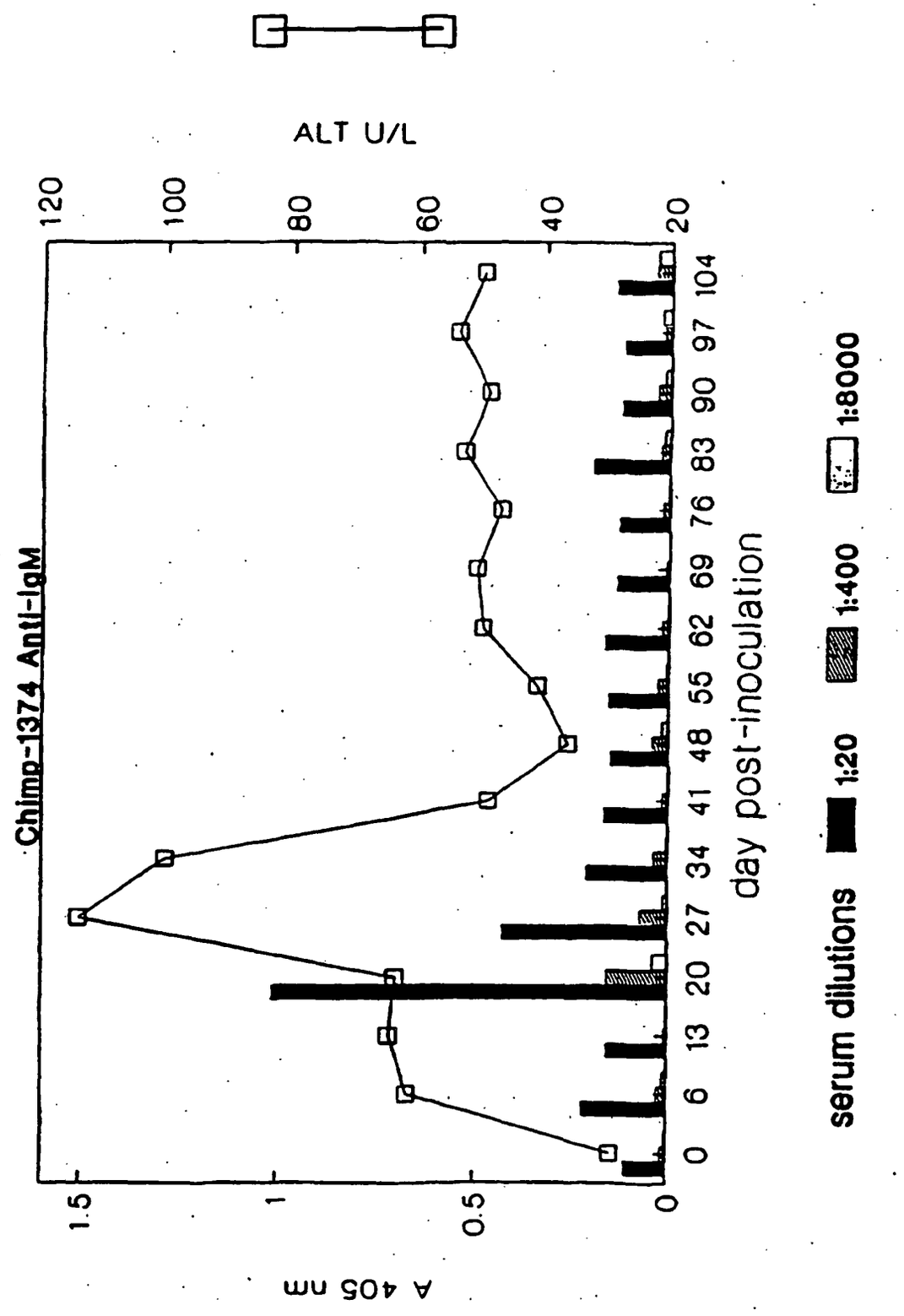

ALT U/L
120
100
80
60
40
20
0
Chimp-1374 Anti-IgG
405 nm
0.8
0.6
0.4
0.2
0
0 6 13 20 27 34 41 48 55 62 69 76 83 90 97 104
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 5B


Chimp-1374 Anti-IgM
A 405 nm
0
0.5
1
1.5
ALT U/L
20
40
60
80
100
120
0 6 13 20 27 34 41 48 55 62 69 76 83 90 97 104
day post-inoculation
serum dilutions
1:20
1:400
1:8000

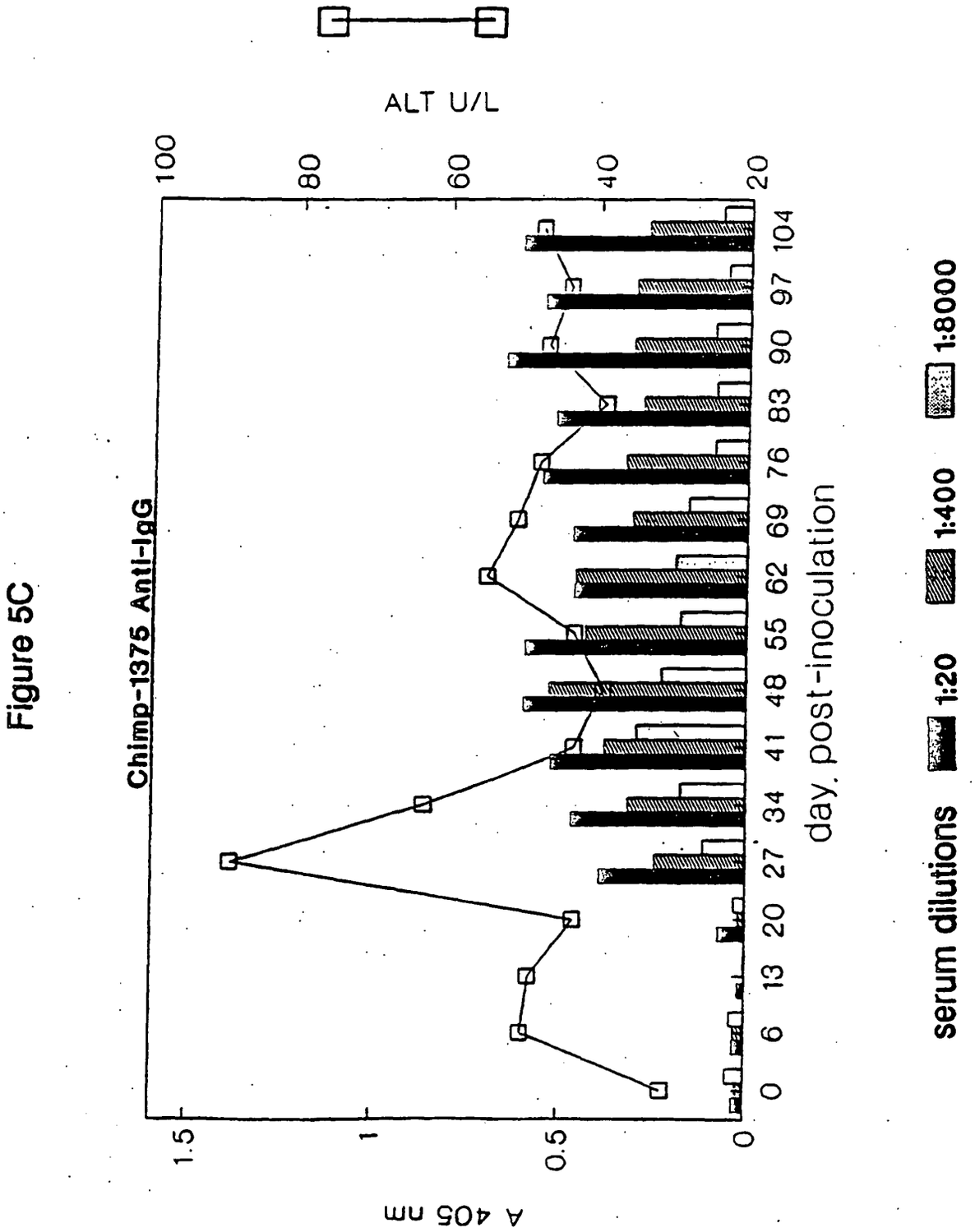

ALT U/L
100
80
60
40
20
Chimp-1375 Anti-IgG
A 405 nm
1.5
1
0.5
0
0 6 13 20 27 34 41 48 55 62 69 76 83 90 97 104
day. post-inoculation
serum dilutions
1:20
1:400
1:8000

Figure 5C

FIG. 5D

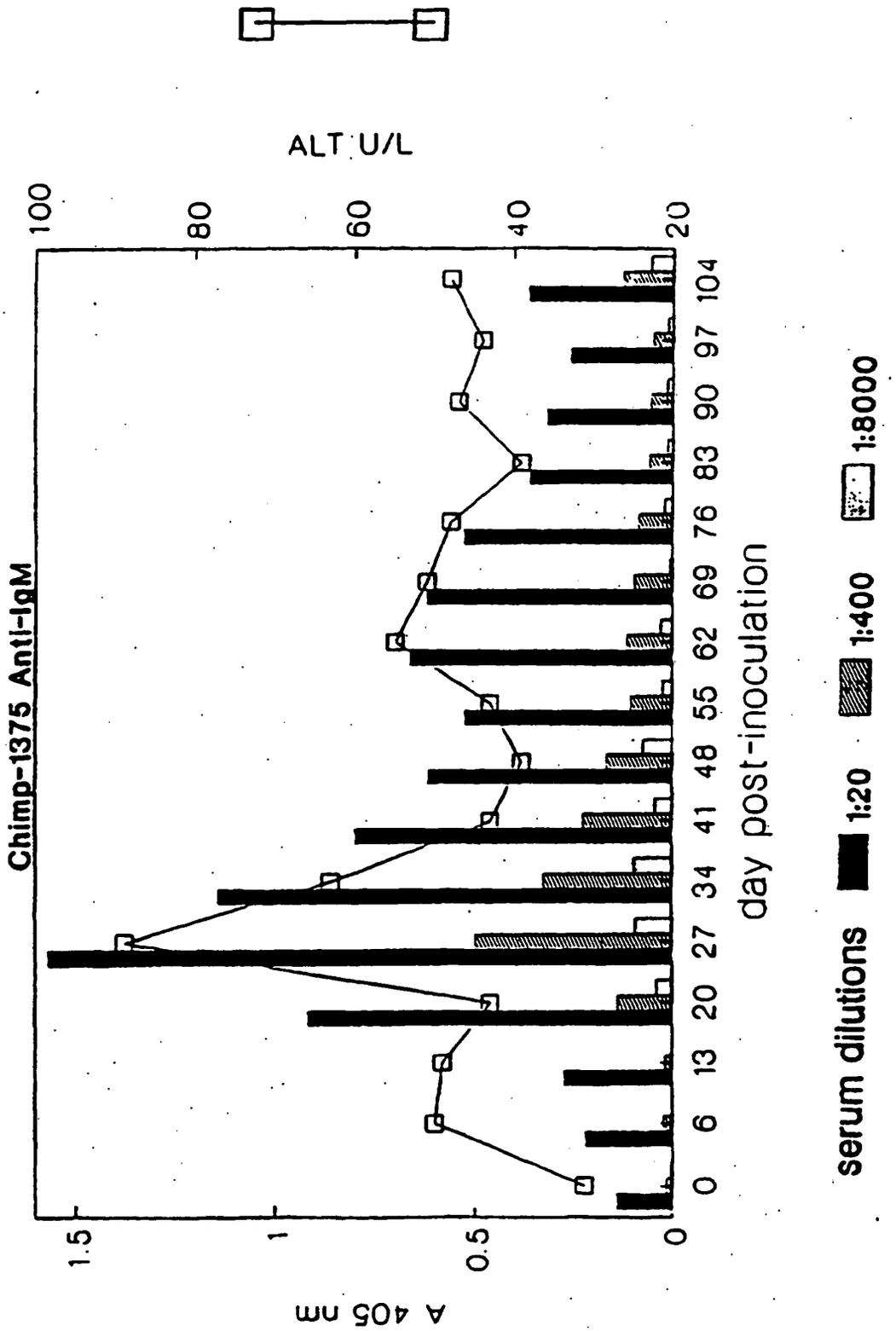

ALT U/L
100
80
60
40
20
Chimp-1375 Anti-IgM
A 405 nm
1.5
1
0.5
0
0 6 13 20 27 34 41 48 55 62 69 76 83 90 97 104
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6A

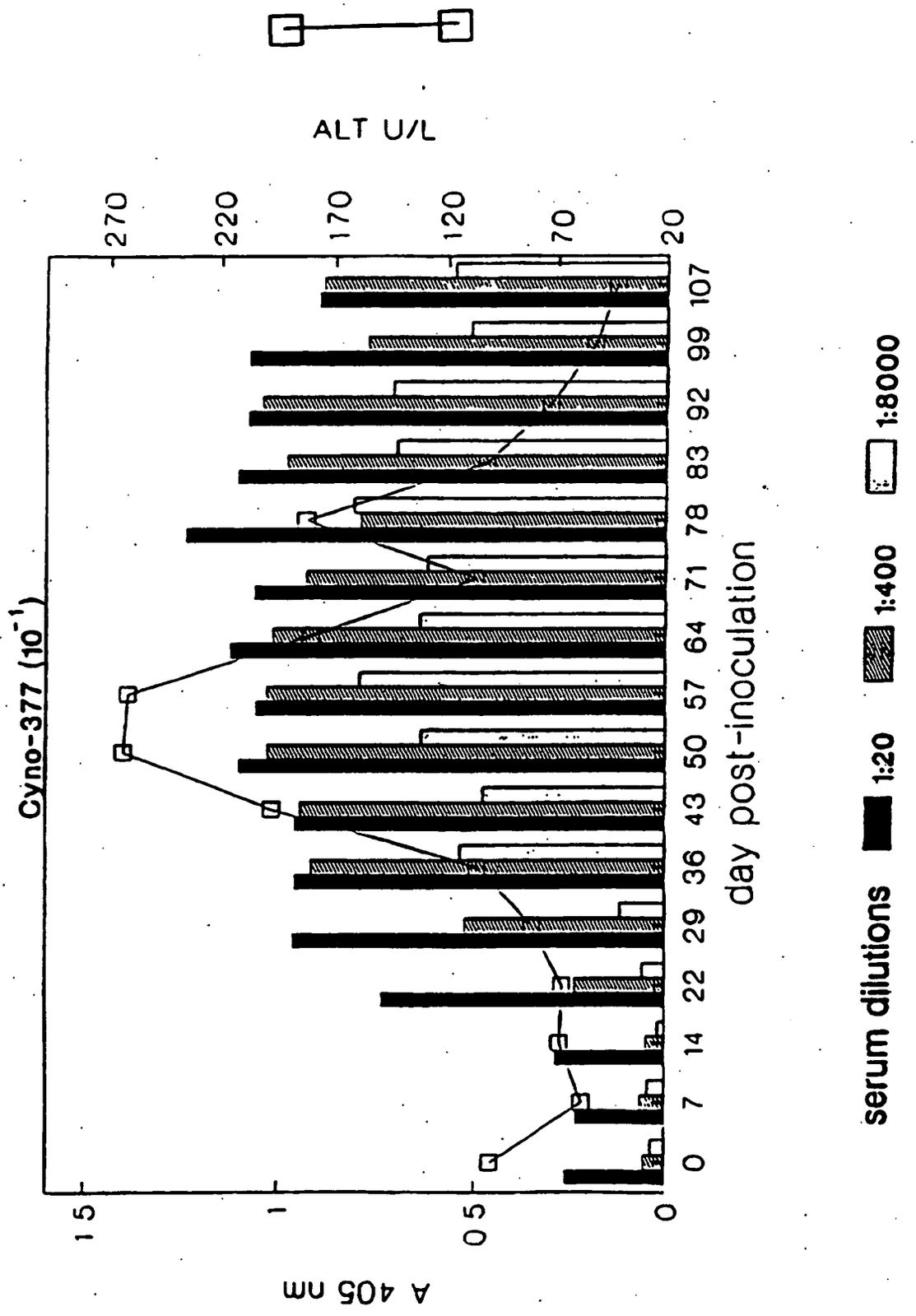

ALT U/L
270
220
170
120
70
20
Cyno-377 (10^{-1})
A 405 nm
1.5
1
0.5
0
0 7 14 22 29 36 43 50 57 64 71 78 83 92 99 107
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6B

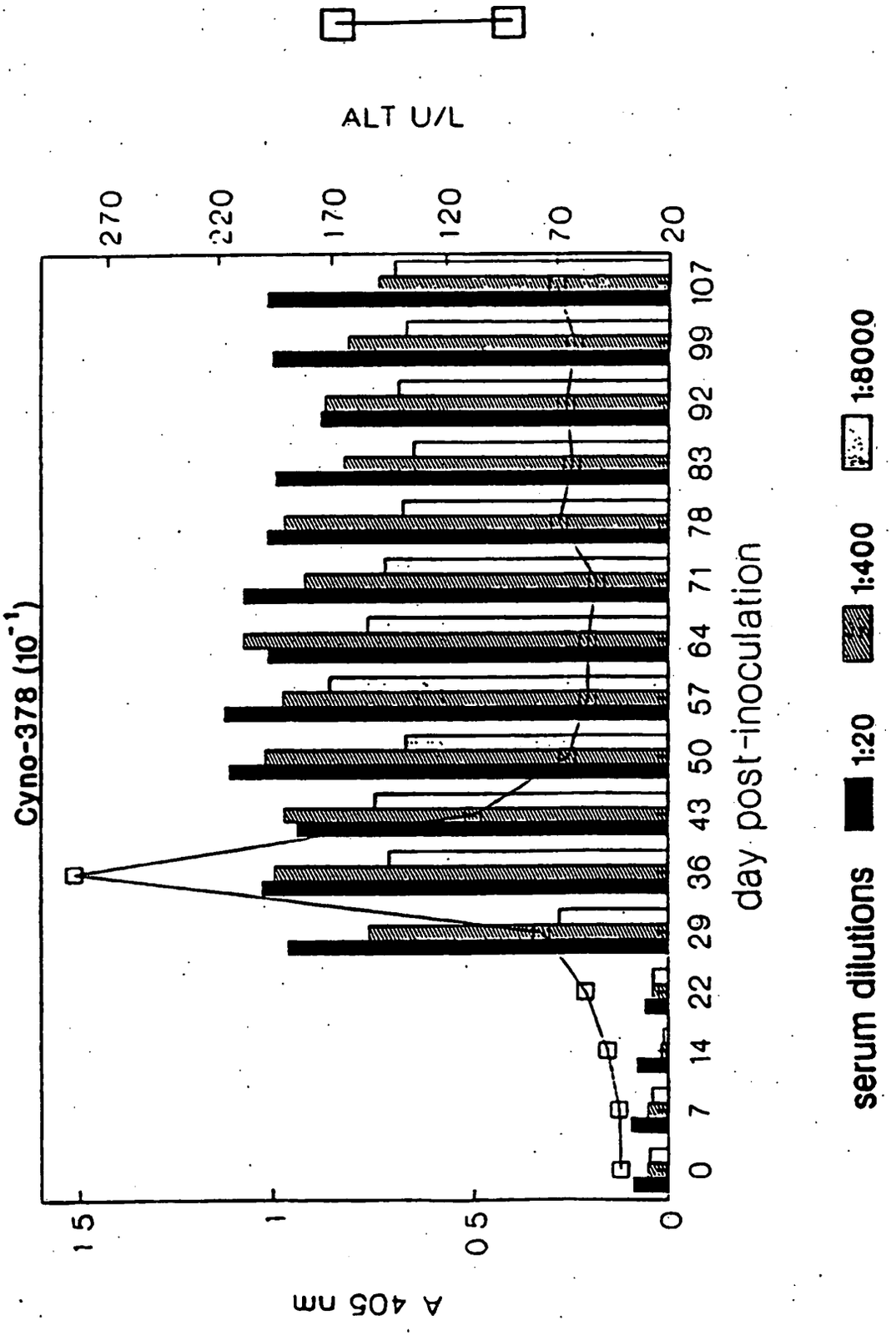

ALT U/L
270
220
170
120
70
20
Cyno-378 (10^{-1})
A 405 nm
1.5
1
0.5
0
0 7 14 22 29 36 43 50 57 64 71 78 83 92 99 107
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6C

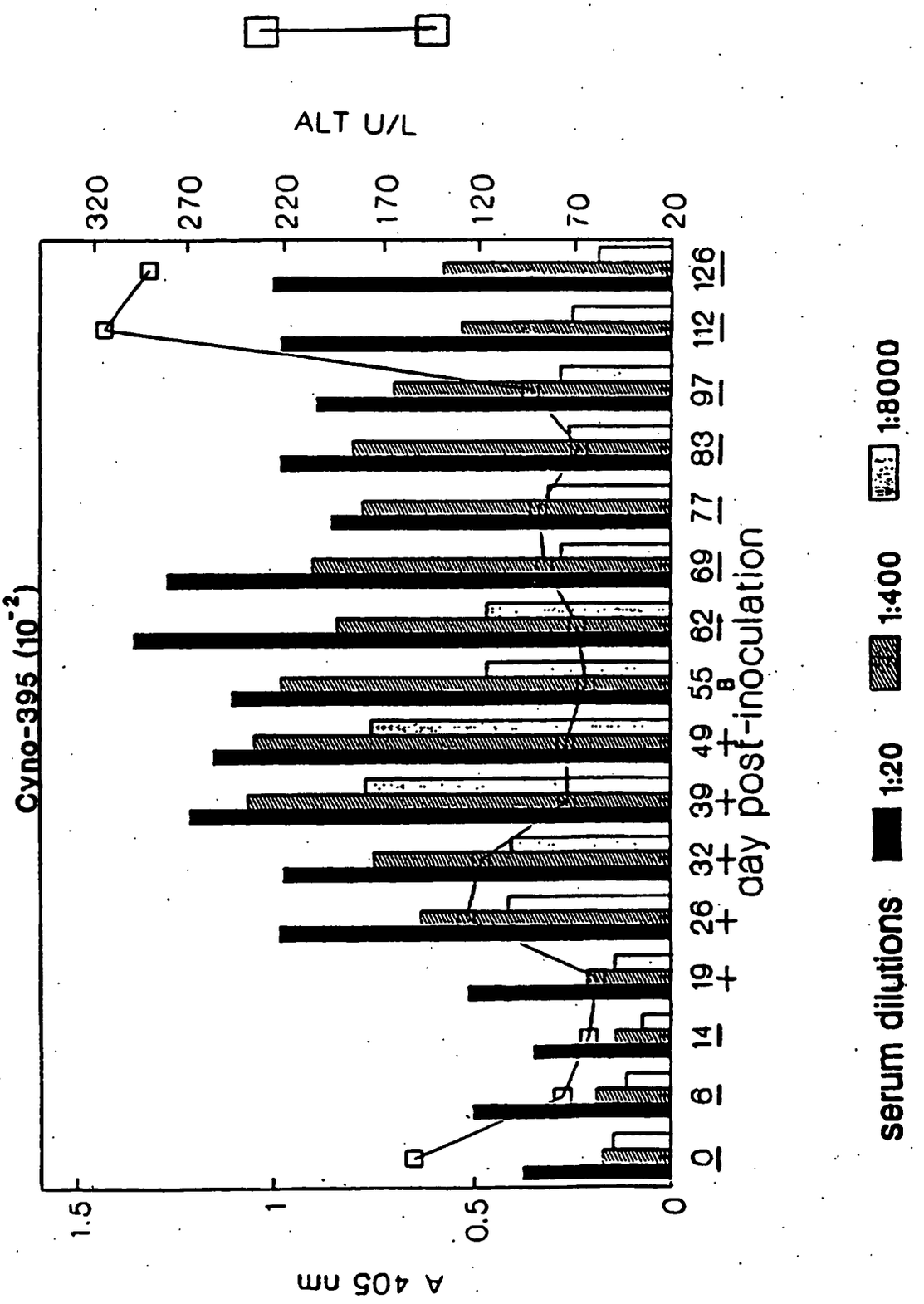

ALT U/L
320
270
220
170
120
70
20
Cyno-395 (10^{-2})
A 405 nm
1.5
1
0.5
0
0
6
14
19
26
32
39
49
55
62
69
77
83
97
112
126
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6D

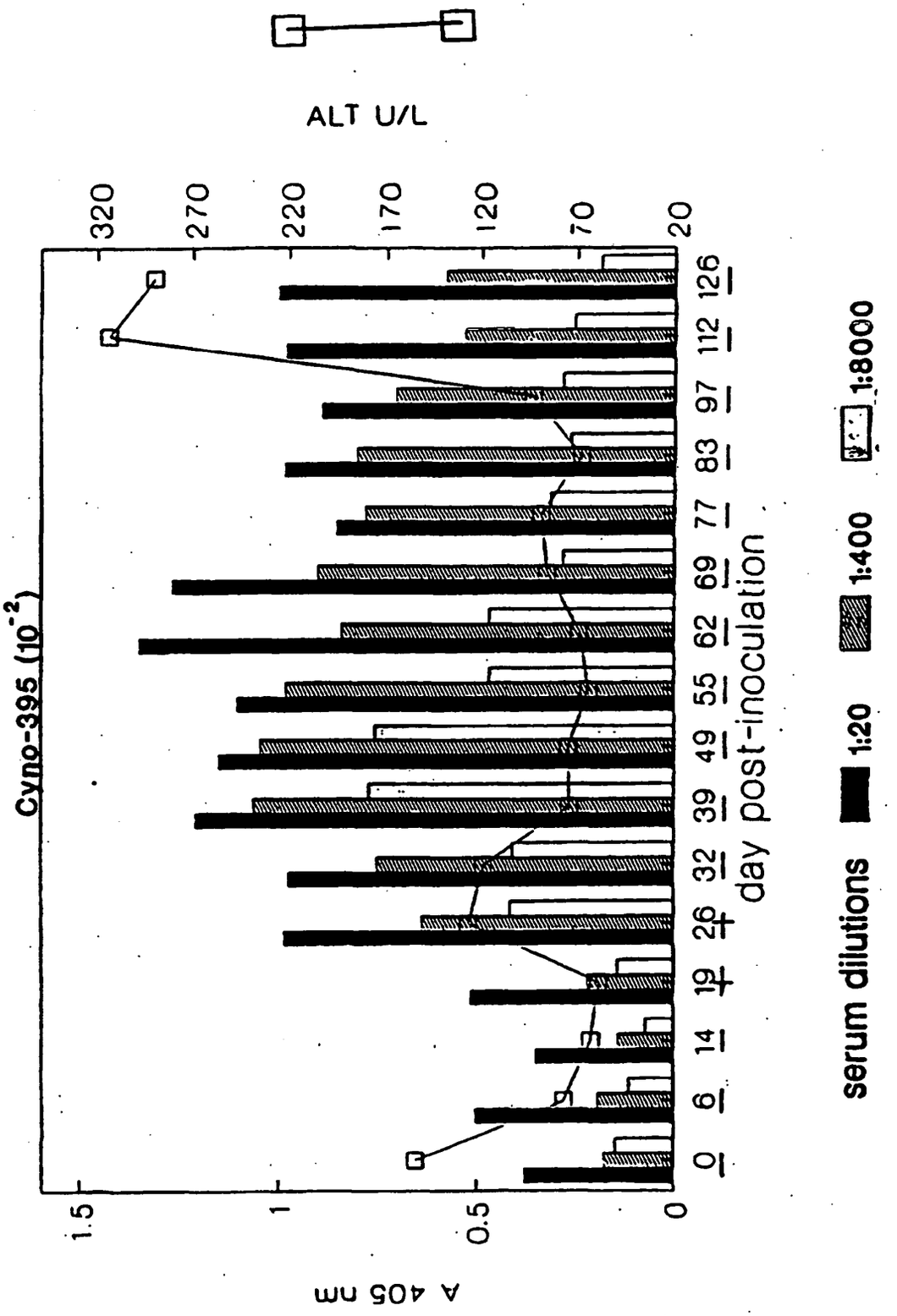

ALT U/L
320
270
220
170
120
70
20
Cyno-395 (10^{-2})
A 405 nm
1.5
1
0.5
0
0
6
14
19
26
32
39
49
55
62
69
77
83
97
112
126
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6E

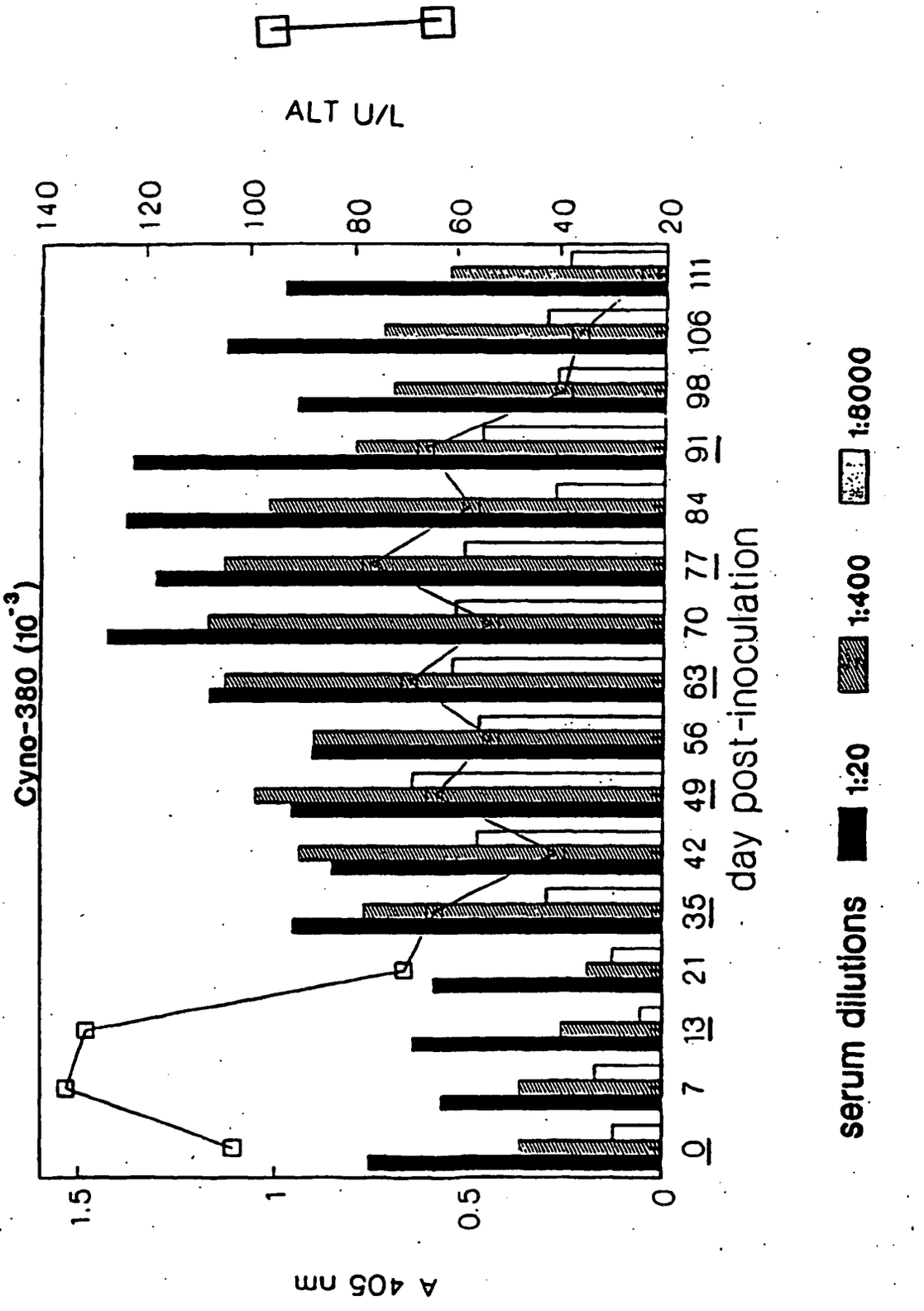

ALT U/L
140
120
100
80
60
40
20
Cyno-380 (10^{-3})
1.5
1
0.5
0
A 405 nm
0
7
13
21
35
42
49
56
63
70
77
84
91
98
106
111
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6F

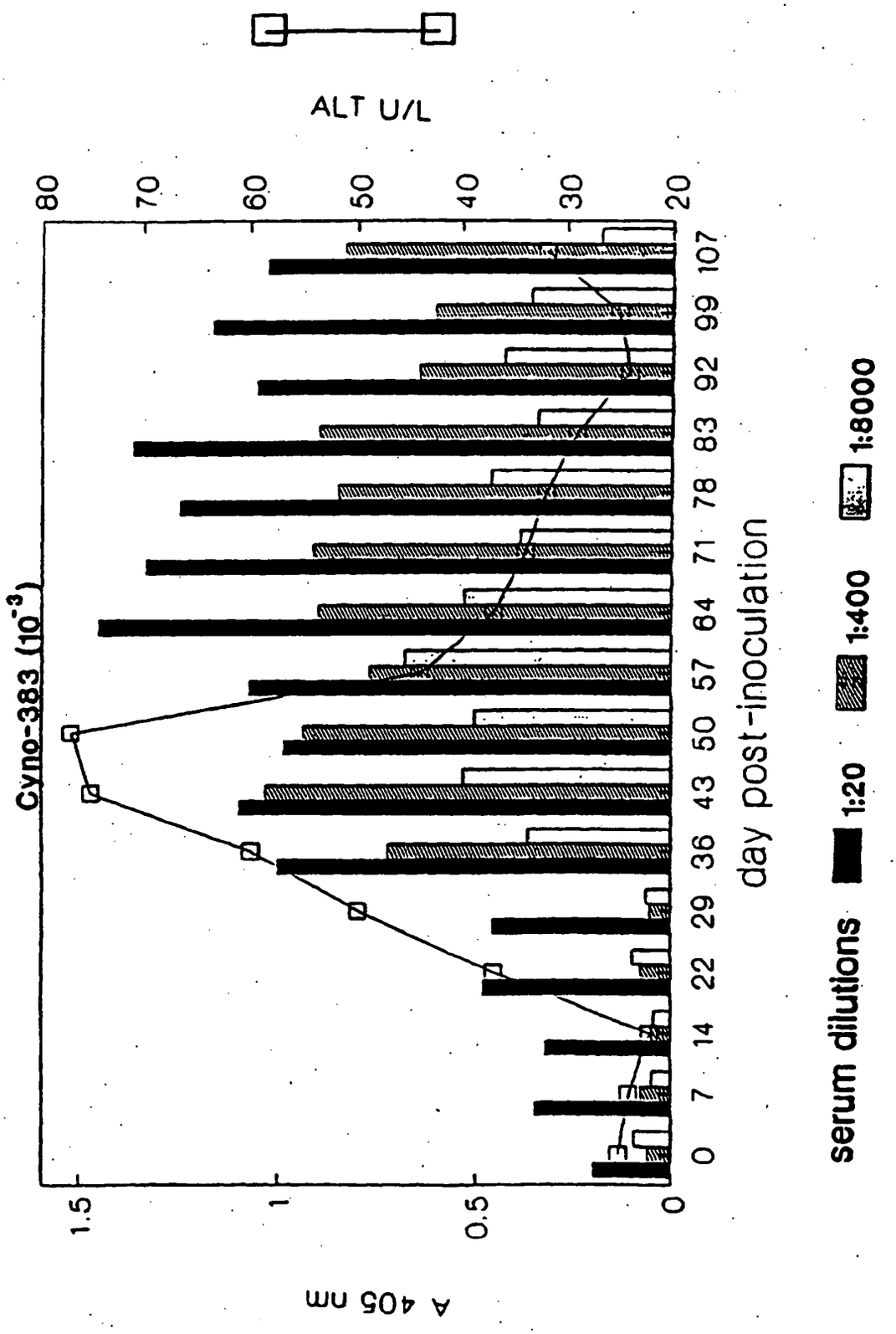

ALT U/L
80
70
60
50
40
30
20
Cyno-383 (10^{-3})
1.5
1
0.5
0
A 405 nm
0
7
14
22
29
36
43
50
57
64
71
78
83
92
99
107
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6G

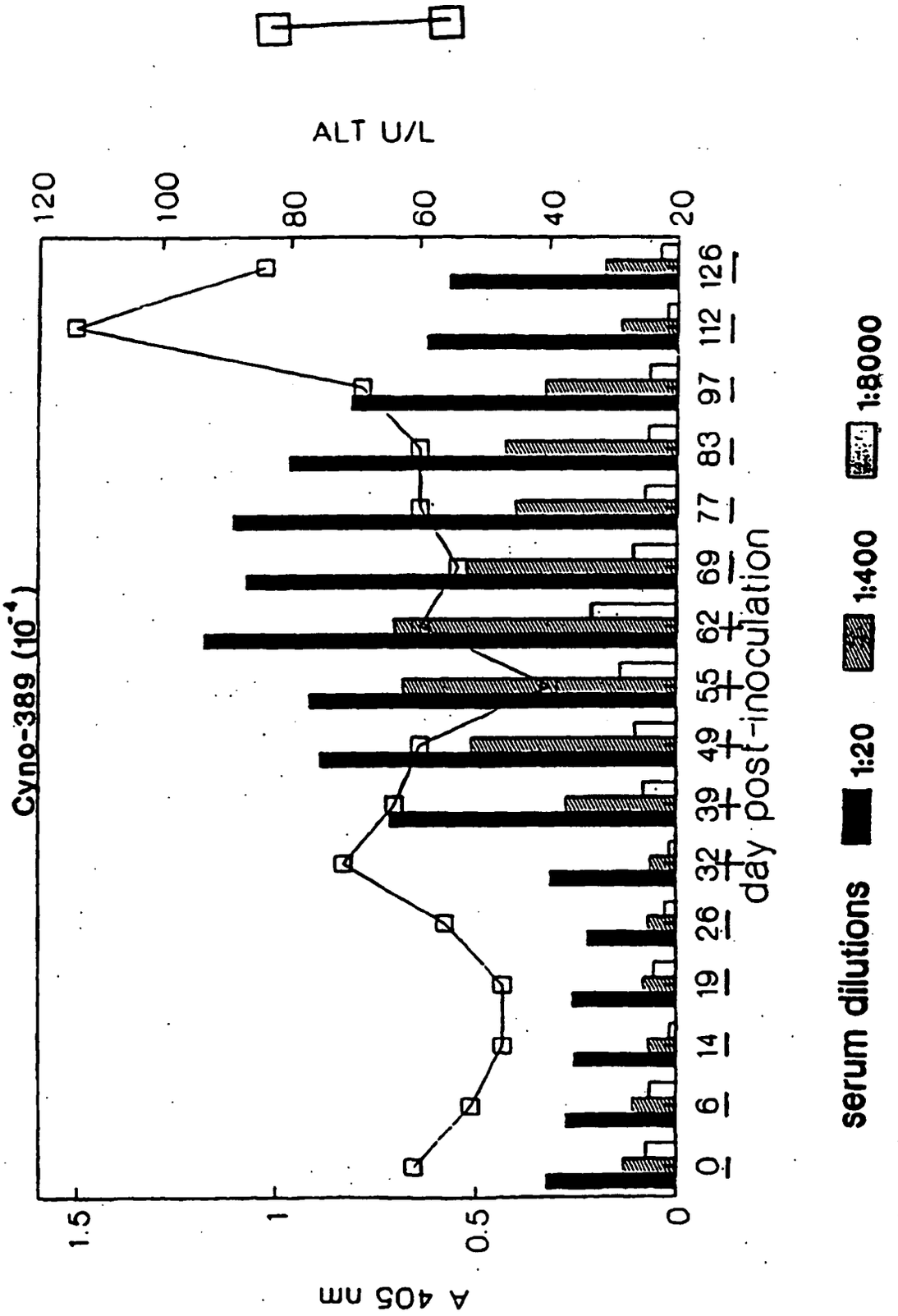

ALT U/L
120
100
80
60
40
20
Cyno-389 (10^{-4})
A 405 nm
1.5
1
0.5
0
0 6 14 19 26 32 39 49 55 62 69 77 83 97 112 126
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6H

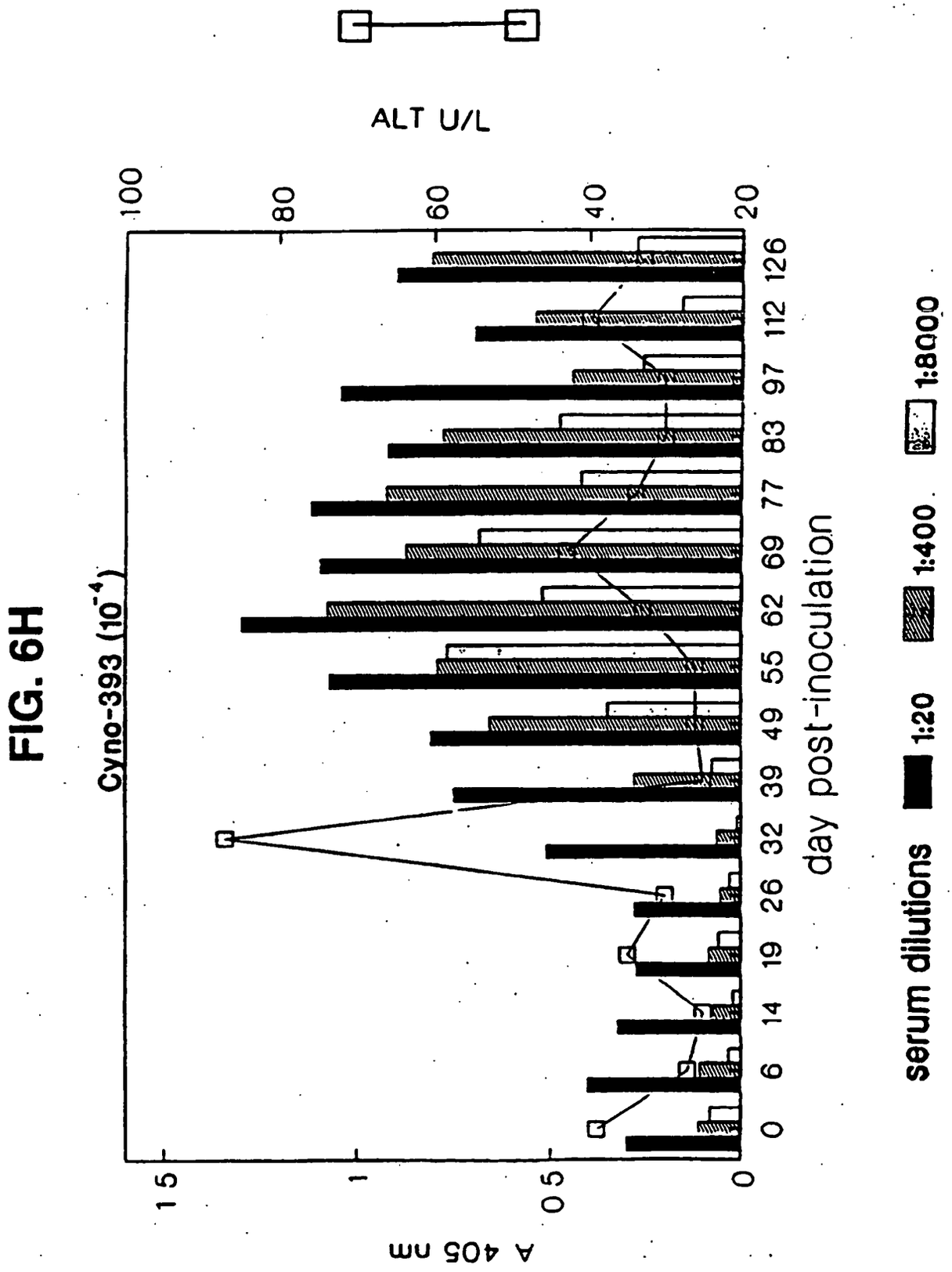

ALT U/L
100
80
60
40
20
Cyno-393 (10^{-4})
A 405 nm
1.5
1
0.5
0
0 6 14 19 26 32 39 49 55 62 69 77 83 97 112 126
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6I

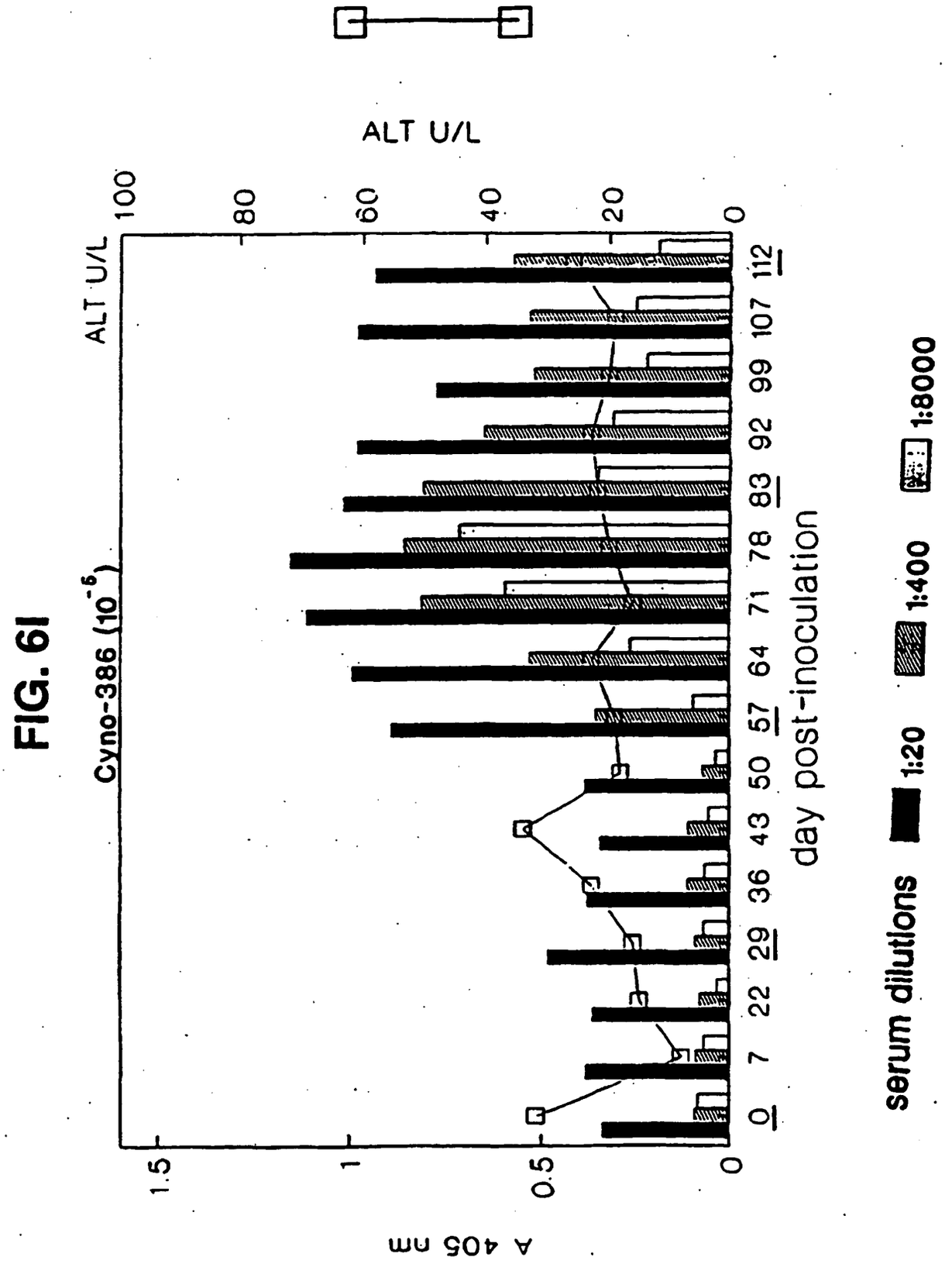
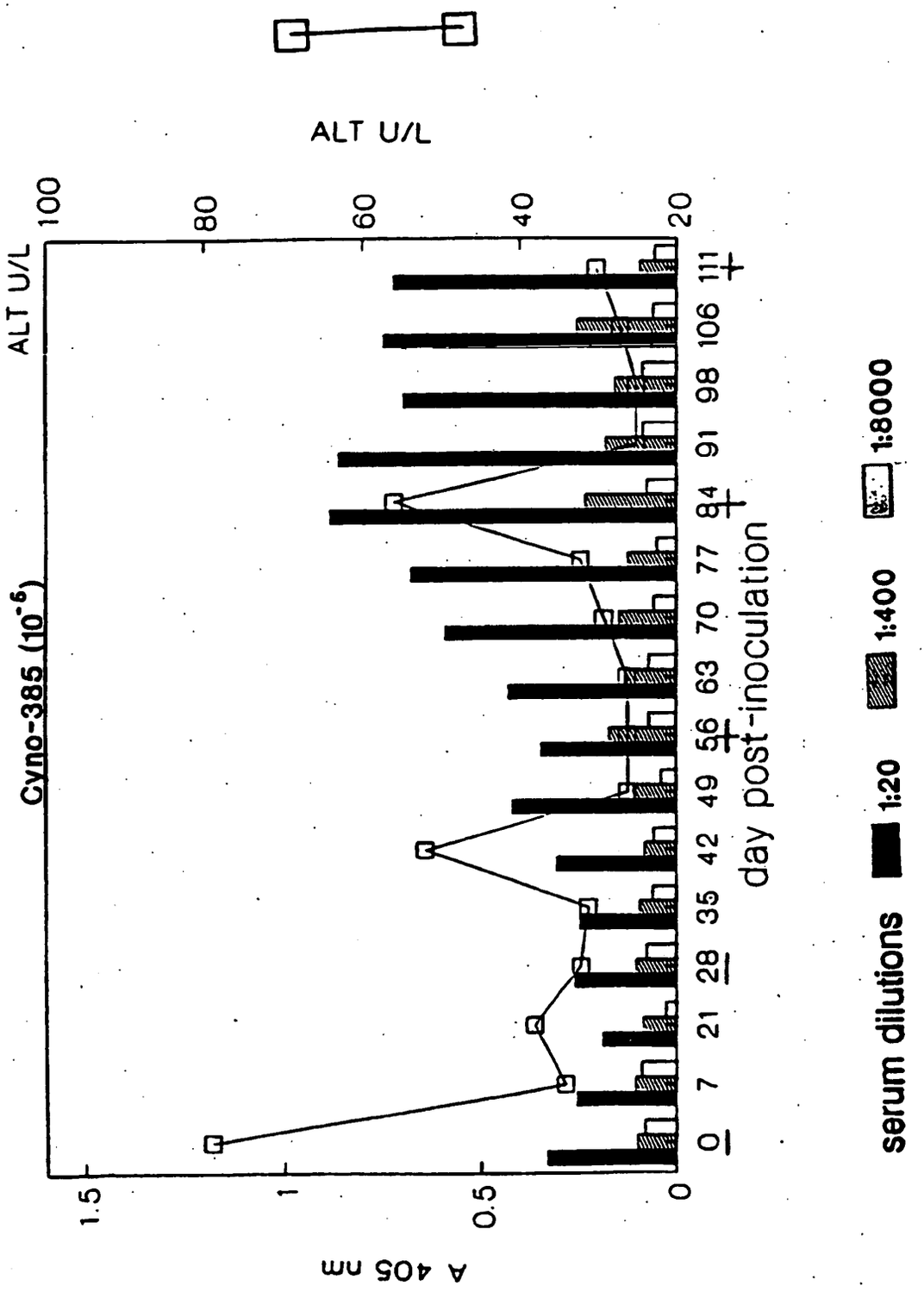

ALT U/L
Cyno-386 (10^{-6})
A 405 nm
day post-inoculation
serum dilutions
1:20
1:400
1:8000

FIG. 6J


ALT U/L
Cyno-385 (10^{-6})
A 405 nm
day post-inoculation
serum dilutions
1:20
1:400
1:8000

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 9115603 A **[0005]**
- US 4683202 A **[0032]**
- WO 184187 A **[0062]**
- EP 171496 A **[0062]**
- EP 173494 A **[0062]**
- WO 8601533 A, Neuberger **[0062]**
- WO 901443 A **[0065]**
- WO 9014424 A **[0065]**
- WO 07947263 A **[0110]**

Non-patent literature cited in the description

- **Balayan, M.S.** Soviet Medical Reviews, Section E, Virology Reviews. Harwood Academic Publishers, 1987, vol. 2, 235-261 **[0003]**
- **Purcell, R.H. et al.** Viral Hepatitis and Liver Disease. Alan R. Liss, 1988, 131-137 **[0003]**
- **Bradley, D.W.** *British Medical Bulletin,* 1990, vol. 46, 442-461 **[0003]**
- **Ticehurst, J.R.** Viral Hepatitis and Liver Disease. Williams and Wilkins, Baltimore, 1991, 501-513 **[0003]**
- **Huang, R.T. et al.** *J. Gen. Virol.,* 1992, vol. 73, 1143-1148 **[0003]**
- **Tam, A.W. et al.** *Virology,* 1991, vol. 185, 120-131 **[0004] [0078]**
- **Tsarev, S.A. et al.** *Proc, Natl, Acad. Sci. USA,* 1992, vol. 89, 559-563 **[0004]**
- **Fry, K.E. et al.** *Virus Genes,* 1992, vol. 6, 173-185 **[0004]**
- **Reyes et al.** *Science,* 1990, vol. 247, 1335-1339 **[0005]**
- **Ichikawa et al.** *Microbiol. Immunol.,* 1991, vol. 35, 535-543 **[0006]**
- **Yarbough et al.** *J. Virology,* 1991, vol. 65, 5790-5797 **[0007]**
- **Purdy et al.** *Archives of Virology,* 1992, vol. 123, 335-349 **[0008]**
- **Favorov et al.** *J. of Medical Virology,* 1992, vol. 36, 246-250 **[0008]**
- **Neil, J.D. et al.** *J. Virol.,* 1991, vol. 65, 5440-5447 **[0009]**
- **Carter, M.J. et al.** *J. Arch. Virol.,* 1992, vol. 122, 223-235 **[0009]**
- **Dawson, G.J. et al.** *J. Virol. Meth,* 1992, vol. 38, 175-186 **[0009]**
- **Lemaitre, M. et al.** *Proc Natl Acad Sci USA,* 1989, vol. 84, 648-652 **[0033]**
- Methods in Immunodiagnosis. John Wiley and Sons, 1980 **[0046]**
- **Campbell et al.** Methods of Immunology. W.A. Benjamin, Inc, 1964 **[0046]**
- **Oellerich, M.** *J.Clin. Chem. Clin. BioChem.,* 1984, vol. 22, 895-904 **[0046]**
- **Cabilly et al.** *Proc. Natl. Acad. Sci. USA,* 1987, vol. 84, 3439 **[0062]**
- **Nishimura et al.** *Canc. Res.,* 1987, vol. 47, 999 **[0062]**
- **Wood et al.** *Nature,* 1985, vol. 314, 446 **[0062]**
- **Shaw et al.** *J. Natl. Cancer Inst.,* 1988, vol. 80, 15553 **[0062]**
- **Morrison S.** *Science,* 1985, vol. 229, 1202 **[0063]**
- **oi et al.** *BioTechniques,* 1986, vol. 4, 214 **[0063]**
- **Jones et al.** *Nature,* 1986, vol. 321, 552 **[0064]**
- **Verhoeyan et al.** *Science,* 1988, vol. 239, 1534 **[0064]**
- **Biedler et al.** *J. Immunol.,* 1988, vol. 141, 4053 **[0064]**
- **Huse et al.** *Science,* 1989, vol. 246, 1275-1281 **[0065]**
- **Goding, J.W.** Monoclonal Antibodies: Principles and Practice. Pladermic Press, Inc, 1983, 56-97 **[0071]**
- **Tsarev, S.A. et al.** *Proc. Natl. Acad. Sci USA,* 1992, vol. 89, 559-563 **[0075]**
- **Tsarev, S.A. et al.** *J. Infect. Dis.,* 1992 **[0075] [0097] [0102]**
- **Winship, P.R.** *Nucleic Acids Rev.,* 1984, vol. 17, 1266 **[0084]**
- **Devereaux, J. et al.** *Nucleic Acids Rev.,* 1984, vol. 12, 387-395 **[0085]**
- **Tsarev, S.A. et al.** *Proc. Natl. Acad, Sci. USA,* 1992, vol. 89, 559-563 **[0086]**
- **Tsarev, S.A. et al.** *Proc. Natl. Acad. Sci, USA,* 1992, vol. 89, 559-563 **[0086]**
- **Laemmli, U.K.** *Nature,* 1970, vol. 227, 680 **[0088]**
- **Ticehurst, J. et al.** *J. Infect. Dis.,* 1992, vol. 165, 835-845 **[0093]**
- **Garci, P. et al.** *J. Infect. Dis.,* 1992, vol. 165, 1006-1011 **[0094]**
- **Farci, P. et al.** *Science,* 1992 **[0094]**
- **Ponzetto, A. et al.** *J Infect. Dis.,* 1987, vol. 155, 72-77 **[0094]**

- **Rizzetto.** *Hepatology,* 1981, vol. 1, 567-574 **[0094]**
- **Tsarev, S.A. et al.** *Proc. Natl. Acad, Sci USA,* 1992, vol. 89, 559-563 **[0102]**
- **Tsarev, SA et al.** *Proc. Nat. Acad. Sci. USA,* 1992, vol. 89, 559-563 **[0102]**
- **Tsarev et al.** *J. Infect. Dis.,* 1992 **[0102]**